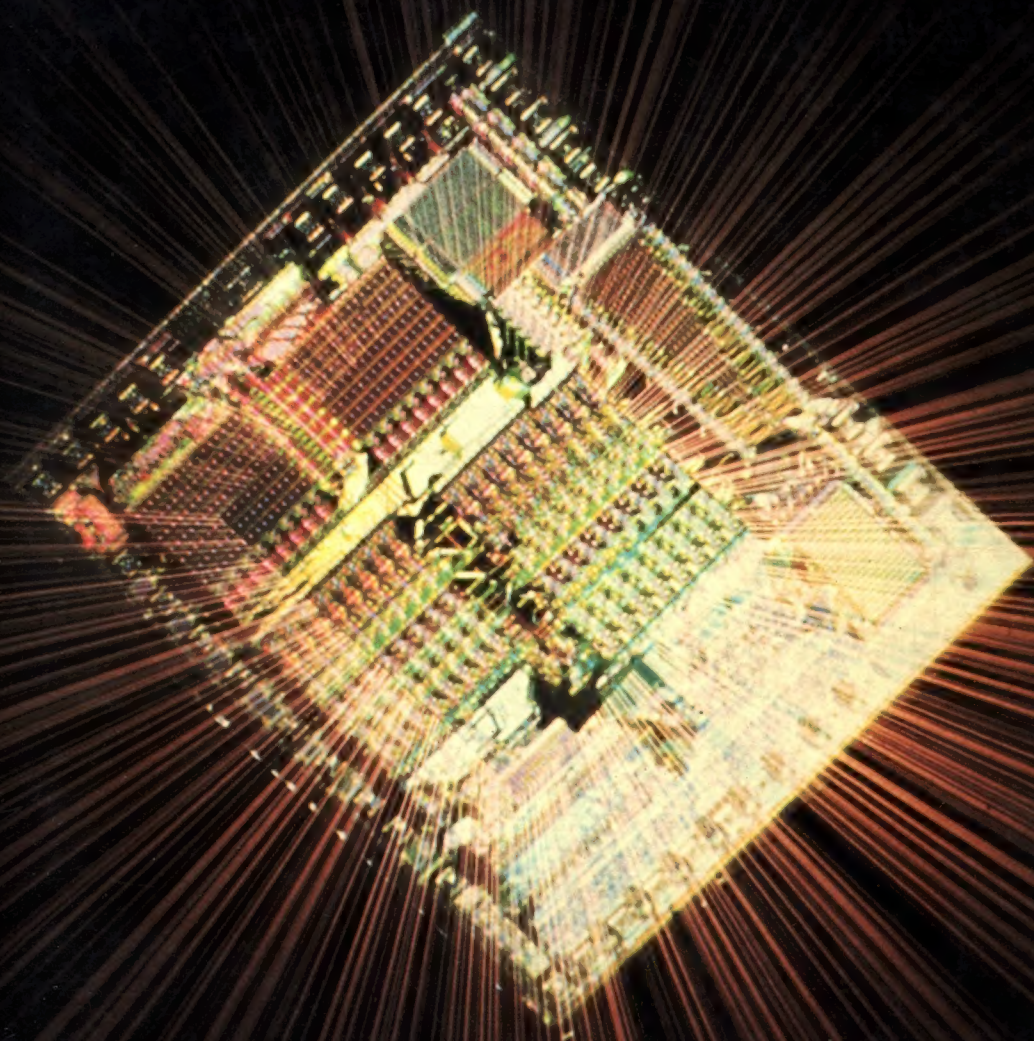


# **Electronics Sourcebook**

## **FOR TECHNICIANS AND ENGINEERS**



**Milton Kaufman   Arthur H. Seidman**



# **ELECTRONICS SOURCEBOOK**

## Other McGraw-Hill Reference Books of Interest

---

### Handbooks

*Avalone and Baumeister* • STANDARD HANDBOOK FOR MECHANICAL ENGINEERS  
*Beeman* • INDUSTRIAL POWER SYSTEMS HANDBOOK  
*Coombs* • BASIC ELECTRONIC INSTRUMENT HANDBOOK  
*Coombs* • PRINTED CIRCUITS HANDBOOK  
*Croft and Summers* • AMERICAN ELECTRICIANS' HANDBOOK  
*Di Giacomo* • VLSI HANDBOOK  
*Fink and Beaty* • STANDARD HANDBOOK FOR ELECTRICAL ENGINEERS  
*Fink and Christiansen* • ELECTRONICS ENGINEERS' HANDBOOK  
*Harper* • HANDBOOK OF ELECTRONIC SYSTEMS DESIGN  
*Harper* • HANDBOOK OF COMPONENTS FOR ELECTRONICS  
*Harper* • HANDBOOK OF THICK FILM HYBRID MICROELECTRONICS  
*Harper* • HANDBOOK OF WIRING, CABLING, AND INTERCONNECTING FOR ELECTRONICS  
*Hicks* • STANDARD HANDBOOK OF ENGINEERING CALCULATIONS  
*Inglis* • ELECTRONIC COMMUNICATIONS HANDBOOK  
*Johnson and Jasik* • ANTENNA ENGINEERING HANDBOOK  
*Juran* • QUALITY CONTROL HANDBOOK  
*Kaufman and Seidman* • HANDBOOK OF ELECTRONICS CALCULATIONS  
*Kaufman and Seidman* • HANDBOOK FOR ELECTRONICS ENGINEERING TECHNICIANS  
*Kurtz* • HANDBOOK OF ENGINEERING ECONOMICS  
*Perry* • ENGINEERING MANUAL  
*Stout* • HANDBOOK OF MICROPROCESSOR DESIGN AND APPLICATIONS  
*Stout and Kaufman* • HANDBOOK OF MICROCIRCUIT DESIGN AND APPLICATION  
*Stout and Kaufman* • HANDBOOK OF OPERATIONAL AMPLIFIER DESIGN  
*Tuma* • ENGINEERING MATHEMATICS HANDBOOK  
*Williams* • DESIGNER'S HANDBOOK OF INTEGRATED CIRCUITS  
*Williams and Taylor* • ELECTRONIC FILTER DESIGN HANDBOOK

### Encyclopedias

CONCISE ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY  
ENCYCLOPEDIA OF ELECTRONICS AND COMPUTERS  
ENCYCLOPEDIA OF ENGINEERING

### Dictionaries

DICTIONARY OF SCIENTIFIC AND TECHNICAL TERMS  
DICTIONARY OF COMPUTERS  
DICTIONARY OF ELECTRICAL AND ELECTRONIC ENGINEERING  
DICTIONARY OF ENGINEERING  
*Markus* • ELECTRONICS DICTIONARY



# **ELECTRONICS SOURCEBOOK**

**For Technicians and Engineers**

**Editors in Chief**

**MILTON KAUFMAN**

*President*

*Electronic Writers and Editors, Inc.*

**ARTHUR H. SEIDMAN**

*(Retired) Professor of Electrical Engineering*

*Acting Dean*

*Pratt Institute, School of Engineering*

**PERRY J. SHENEMAN**

*Associate Editor*

**McGRAW-HILL BOOK COMPANY**

*New York St. Louis San Francisco Auckland Bogotá  
Caracas Colorado Springs Hamburg Lisbon  
London Madrid Mexico Milan Montreal  
New Delhi Oklahoma City Panama Paris  
San Juan Sao Paulo Singapore  
Sydney Tokyo Toronto*

Library of Congress Cataloging-in-Publication Data

Electronics sourcebook for technicians and engineers / editors in chief, Milton Kaufman, Arthur H. Seidman ; Perry J. Sheneman, associate editor.

p. cm.

Rev. ed. of: Handbook for electronics engineering technicians, 2nd ed, c1984.

Bibliography: p.

Includes index.

ISBN 0-07-033559-1

I. Electronics—Handbooks, manuals, etc. I. Kaufman, Milton.

II. Seidman, Arthur H. III. Sheneman, Perry J. IV. Handbook for electronics engineering technicians.

TK7825.E38 1988

621.381—dc19

87-33626

CIP

Copyright © 1988 by McGraw-Hill, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

ISBN 0-07-033559-1

The material in this volume is condensed from *Handbook for Electronics Engineering Technicians*, Second Edition, edited by Milton Kaufman and Arthur H. Seidman.

Copyright © 1984, 1976 by McGraw-Hill, Inc.

All rights reserved.

Printed and bound by R. R. Donnelley and Sons Company.

# Contents

List of Contributors

xxiii

Preface

xxv

## 1. CHARACTERISTICS OF RESISTORS

1.1	Introduction	1-1
1.2	General Description	1-1
1.3	Resistor Terms and Parameters	1-3
	Resistance	1-3
	Tolerance	1-3
	Temperature Coefficient of Resistance	1-3
	Power Rating	1-3
	Rated Continuous Working Voltage	1-4
	Critical Resistance	1-4
	Noise	1-4
	Frequency Effects	1-4
	Drift	1-5
	Taper	1-5
	Resolution	1-6
	End Resistance	1-6
	Contact Resistance	1-6
	Wiper (Slider) Current	1-6
	Setting Stability	1-6
1.4	Fixed Resistors	1-8
	Carbon Composition	1-8
	Metal Film	1-8
	Thin film	1-8
	Tin oxide	1-8
	Metal glaze	1-8
	Cermet	1-9
	Bulk film	1-9
	Carbon Film	1-9
	Wire-wound	1-10
1.5	Color Coding of Resistors	1-11
1.6	Connecting Resistors	1-11
	Resistors in Series	1-11
	Resistors in Parallel	1-12

1.7	Special Resistors	1-13
	High-Voltage Resistors	1-13
	High-Megohm Resistors	1-13
	DIP Networks	1-13
1.8	Variable Resistors (Pots)	1-13
	Types of Materials	1-15
	Carbon	1-15
	Cermet	1-15
	Conductive plastic	1-15
	Wire-wound	1-16
	Rheostat	1-16
1.9	Chip Resistors	1-16
1.10	Thermistors	1-16
	Characteristic Curves and Parameters	1-16
	Resistance-temperature characteristic	1-16
	Static voltampere characteristics	1-17
	Current-time characteristic	1-18
1.11	Varistors	1-19

## 2. CHARACTERISTICS OF CAPACITORS

2.1	Introduction	2-1
	Applications	2-1
2.2	General Description	2-1
	Relative Dielectric Constant	2-3
	Charge and Energy	2-3
2.3	Capacitor Terms and Parameters	2-4
	Capacitance	2-4
	Ambient Temperature	2-4
	Tolerance	2-4
	Temperature Coefficient	2-5
	Working Voltage	2-5
	Breakdown Voltage	2-5
	DC Leakage	2-5
	Insulation Resistance	2-5
	Capacitive Reactance	2-5
	Power Factor	2-5
	Equivalent Series Resistance	2-6
	Impedance	2-6
	Dissipation Factor	2-6
	Quality Factor ( $Q$ )	2-7
	Ripple Current and Voltage	2-7
2.4	Fixed Capacitors	2-7
	Mica	2-7
	Silvered Mica	2-8
	Paper	2-8
	Plastic Film	2-8
	Ceramic	2-8
	Glass	2-8
	Vacuum	2-9
	Energy Storage	2-9
	Metallized Dielectrics	2-9
	Electrolytic	2-9



2.5	Connecting Capacitors	2-10
	Capacitors in Parallel	2-10
	Capacitors in Series	2-10
2.6	Color Coding of Capacitors	2-11
2.7	Variable Capacitors	2-11
	Trimmers	2-12
2.8	Chip Capacitors	2-13
2.9	Voltage-Variable Capacitors (Varactors)	2-15
	Parameters	2-16
	Total diode capacitance ( $C_t$ )	2-16
	Series resistance ( $R_s$ )	2-16
	Figure of merit ( $Q$ )	2-16
	Cutoff frequency ( $f_{co}$ )	2-16
	Variation of capacitance with temperature	2-16
	$C_t$ ratio	2-16
	Conversion efficiency ( $\eta$ )	2-17

### 3. COILS

3.1	Introduction	3-1
3.2	Types of Ferrous Cores and Their Effects in Coils	3-1
	Iron-Power Cores	3-1
	Ferrite Cores	3-1
	Laminated Cores	3-1
	Tuning Cores, Slugs, and Toroids	3-3
	Core Permeability	3-3
3.3	Types of Nonferrous Cores	3-3
3.4	Computing the Inductance of Air-Core Coils	3-3
3.5	Mutual Inductance	3-5
	Coupled Coils in Series and in Parallel	3-6
	Computing the Mutual Inductance of Single-Layer Air-Core Coils	3-7
3.6	Computing the Value of Inductance from Measurements	3-9
3.7	The Coefficient of Coupling	3-10
3.8	Coil and Conductor Resistance at Radio Frequencies	3-10
	Eddy Currents in Conductors	3-10
	Dielectric Losses	3-10
	Distributed Capacitance	3-11

### 4. MAGNETIC CIRCUITS

4.1	Introduction	4-1
4.2	Magnets, Natural and Artificial	4-1
4.3	Magnetic Materials	4-1
4.4	The Magnetic Compass	4-3
4.5	Magnetic Poles	4-3
	Coulomb's Law	4-4
	The Unit Magnetic Pole	4-4
	Magnetic Pole Computations	4-4
4.6	The Earth's Magnetic Field	4-5
4.7	Electromagnetism	4-5
	The Right-Hand Rule and the Left-Hand Rule for Magnetic Fields Around Wires	4-6
4.8	The Right-Hand Rule and the Left-Hand Rule for Coils	4-6

4.9	Units of Measurement in Magnetics	4-7
	Magnetic Flux	4-7
	Unit Poles	4-7
	Flux Density	4-8
	Magnetomotive Force	4-8
	Reluctance	4-9
	Comparison of Electric and Magnetic Circuits	4-9
	Magnetizing Force	4-10
	Relative Permeability	4-11

## 5. TRANSFORMERS

5.1	Introduction	5-1
5.2	The Basic Transformer	5-1
5.3	Mutual Induction	5-1
5.4	Transformer Voltage Relationships	5-4
5.5	Transformer Current Relationships	5-5
5.6	The Fundamental Equation of the Transformer	5-6
5.7	Discussion of No-Load Conditions	5-8
5.8	Vector Representation of No-Load Conditions	5-9
5.9	Conditions Under Load	5-11
5.10	Operating a Transformer on Lower or Higher than Rated Frequency	5-12
5.11	Connecting Transformers in Series and Parallel	5-14
	Series Operation	5-14
	Parallel Operation	5-14
	Phasing Transformers for Parallel Operation	5-14
5.12	Transformer Losses	5-15
	Copper Loss	5-15
	Determination of Copper loss	5-15
	Core-Loss Test	5-16
5.13	Efficiency	5-17
5.14	Equivalent Resistances	5-18
5.15	Voltage Regulation	5-19
5.16	The Isolation Transformer	5-21
5.17	The Autotransformer	5-21
5.18	The Variable Autotransformer	5-24
	Using Variable Autotransformers at Higher than Rated Voltages	5-26
	Switching	5-26
5.19	Instrument Transformers	5-26
	The Potential Transformer	5-27
	The Current Transformer	5-27

## 6. PRACTICAL CIRCUIT ANALYSIS

6.1	Introduction	6-1
6.2	Notation	6-1
6.3	Definitions	6-1
	Passive and Active Elements	6-1
	Ideal Elements	6-1
	Resistor $R$	6-1
	Inductor $L$	6-2

Capacitor $C$	6-2
Voltage source $e$	6-2
Current source $i$	6-2
Dependent (controlled) sources	6-2
Model	6-2
Linear Element	6-3
Nonlinear Element	6-3
Linear Circuit	6-3
Nonlinear Circuit	6-3
Series Circuit	6-3
Parallel Circuit	6-3
6.4 $RC$ and $RL$ Circuits	6-3
6.5 Energy and Power	6-5
6.6 Circuit Laws and Theorems	6-5
Kirchhoff's Laws	6-5
Conductance	6-8

## 7. METERS AND MEASUREMENTS

7.1 Introduction	7-1
7.2 Parallax and Other Problems in Using Instruments	7-1
7.3 Percent Accuracy	7-3
7.4 Basic Instruments and Their Use for Servicing Electronics	
Equipment	7-3
Ammeters	7-3
Voltmeters	7-4
Ohmmeters	7-5
Multimeters	7-6
7.5 Meter Movements in Measuring Instruments	7-8
Meters that Utilize the Magnetic Field of a Current for Measurement	7-9
Moving-Coil Meters	7-9
Meters that Depend on the Heating Effect of a Current	7-11
7.6 Meter Circuits	7-11
Calculation of Meter Shunts	7-11
Calibration of Current-Reading Meters	7-13
Calculation of Meter Multipliers	7-14
7.7 Vacuum-Tube Voltmeters	7-16
7.8 Digital Multimeters	7-18
Digital Displays	7-19
Theory of Operation	7-20
The $\frac{1}{2}$ digit	7-21
Digital meter specifications	7-21

## 8. SEMICONDUCTOR DEVICES AND TRANSISTORS

8.1 Introduction	8-1
8.2 Elementary Semiconductor Theory	8-1
Atomic Structure	8-1
Energy Bands	8-3
Doping	8-4
n-Type Silicon	8-4
p-Type Silicon	8-5

8.3	Junction Diode	8-5
	Diode Characteristics	8-5
	Rectifier, R-F, and Switching Diodes	8-7
	Diode Parameters	8-7
8.4	Zener Diodes	8-8
	Zener and Avalanche Breakdown	8-8
	Regulator and Reference Diodes	8-9
	Double-Anode Regulator	8-9
	Zener Diode Parameters	8-9
8.5	Bipolar Junction Transistor	8-9
	Operation of the BJT	8-10
	BJT Parameters and Curves	8-12
8.6	Field-Effect Transistor	8-13
	Junction FET	8-13
	Operation	8-14
	MOSFET	8-15
	Operation of depletion-type MOSFET	8-15
	Operation of enhancement-type MOSFET	8-17
	FET Parameters	8-17
8.7	Silicon-Controlled Rectifier	8-18
	Operation of the SCR	8-19
	Light-Activated SCR	8-20
	Characteristic Curves	8-20
	SCR Parameters	8-21
8.8	Bidirectional Triode Thyristor	8-21
8.9	Unijunction Transistor	8-22
	Operation of the UJT	8-22
	Programmable Unijunction Transistor	8-24
	UJT Curves and Parameters	8-24
8.10	Light-Emitting Diode	8-25
	Characteristic Curves	8-25
	LED Parameters	8-26
8.11	Other Semiconductor Devices	8-26
	Tunnel Diode	8-26
	Schottky Barrier Diode	8-27
	Photodiodes	8-27
	Phototransistor	8-28
	Opto-Isolator	8-28
	Charge-Coupled Device	8-28
	Microwave Power Diodes	8-29
8.12	Semiconductor Chips	8-30
	Flip Chip	8-30
	Beam-Lead Device	8-30
	Leadless Inverted Device	8-30
8.13	Understanding Data Sheets	8-30

## 9. INTEGRATED CIRCUIT TECHNOLOGY

9.1	Introduction	9-1
9.2	Dimensioning Integrated Circuits	9-1
9.3	Monolithic Technology	9-1
	Basic Processing Steps	9-2
	P-type substrate	9-3



Epitaxial layer	9-3
Oxide layer	9-3
Photoresist	9-4
Artwork	9-4
Exposure of collector mask	9-6
Diffusion	9-6
Etching	9-7
p-type diffusion	9-7
Oxidation and resist	9-8
Exposure to base mask	9-8
Base diffusion	9-8
Oxidation and resist	9-8
Forming the emitter	9-8
Exposure to $n^+$ mask	9-8
$n^+$ diffusion	9-8
Oxidation and resist	9-8
Exposure to contact mask	9-8
Metallization	9-8
Exposure to interconnection mask	9-9
Testing and packaging	9-9
9.4 Electrical Isolation	9-9
Junction Isolation	9-9
Dielectric Isolation	9-9
Beam-Lead Isolation	9-9
Parasitics	9-10
9.5 Properties and Characteristics of Diffused Elements	9-10
Bipolar Junction Transistor (BJT)	9-10
Lateral pnp Transistor	9-10
Field-Effect Transistor (FET)	9-11
Diode	9-11
Resistor	9-11
Four-Point Probe	9-12
Capacitor	9-12
Inductor	9-13
9.6 Making a Monolithic IC	9-13
Isolation mask	9-14
p-diffusion mask	9-14
$n^+$ diffusion mask	9-14
Contact and interconnection masks	9-14
Buried Crossover	9-15
9.7 Large-Scale Integration (LSI)	9-16

## 10. TUNED CIRCUITS

10.1 Introduction	10-1
Analogies	10-2
Some Uses of Tuned Circuits	10-2
10.2 Types of Tuned Circuits	10-3
Fixed-Tuned Circuits	10-3
Tunable Circuits	10-4
Circuit Bandwidth	10-4
Single-Tuned and Double-Tuned Circuits	10-4
Series-Tuned and Parallel-Tuned Circuits	10-5

10.3	Comparison of Series-Tuned and Parallel-Tuned Circuits	10-5
10.4	Series-Tuned Circuits	10-5
	Applications	10-5
	Fundamentals	10-6
	Series-Tuned Voltages at Resonance	10-8
10.5	Parallel-Tuned Circuits	10-10
	Applications	10-10
	Fundamentals	10-11
10.6	Selectivity Bandwidth, and Q Factor	10-15
	Bandwidth	10-15
	Q Factor	10-16
	Application Requirements	10-18

## 11. FILTERS

11.1	Introduction	11-1
11.2	Classification of Filters	11-1
	Applications	11-2
11.3	Filter Terminology	11-4
11.4	Filter Components	11-6
11.5	Methods of Designing LC Filters	11-6
	Image-Parameter Design	11-6
	Network Method of Design	11-7
11.6	The Significance of Filter Impedance	11-7
11.7	Design of Constant- $k$ Filters	11-9
	Low-Pass Filter Design	11-9
	High-Pass Constant- $k$ Filter Design	11-11
	Bandpass Filter Design	11-12
	Band-Rejection Filter Design	11-14
11.8	Summary of Constant- $k$ Filter Characteristics	11-15
11.9	$m$ -Derived Filters	11-15
	Low-Pass $m$ -Derived Filter Design (Series Type)	11-17
	Low-Pass $m$ -Derived Filter Design (Shunt Type)	11-17
	High-Pass $m$ -Derived Filter Design (Series Type)	11-18
	High-Pass $m$ -Derived Filter Design (Shunt Type)	11-19
11.10	Summary and Design Example of $m$ -Derived Filters	11-19
11.11	Some Additional Types of Filters	11-21
	Butterworth Filter	11-22
	Chebyshev Filter	11-22
	Crystal Filter (sometimes called "Quartz Filter")	11-22
	Ceramic Filter	11-23
	Mechanical Filters (also called "Magnetostrictive Filters")	11-25
11.12	Active Filters	11-26
	High-pass filter	11-28
	Bandpass filter	11-29
	Band-reject filter	11-29

## 12. TRANSISTOR AMPLIFIERS AND OSCILLATORS

12.1	Introduction	12-1
12.2	Classification of Amplifiers	12-1
12.3	Characteristics of Amplifiers	12-2

Voltage Gain	12-3
Current Gain	12-3
Power Gain	12-3
Input Resistance	12-3
Output Resistance	12-3
Bandwidth	12-3
Distortion	12-4
Harmonic distortion	12-4
Intermodulation distortion	12-4
Slewing Rate	12-4
12.4 Biasing and Stabilization	12-4
Biasing the Common-Emitter Amplifier	12-4
Biasing the Common-Base Amplifier	12-6
Biasing the Emitter Follower	12-7
Stabilizing the Common-Emitter Amplifier	12-7
Current-feedback stabilization	12-8
Voltage-feedback stabilization	12-8
Bypass and decoupling capacitors	12-9
Diode stabilization	12-9
Biasing JFET and Depletion-Type MOSFET Amplifiers	12-9
Biasing the Enhancement-Type MOSFET	12-11
12.5 Small-Signal Amplifiers	12-11
Analyzing the BJT Amplifier	12-11
Simplified hybrid model	12-12
Application of the simplified hybrid model	12-13
Practical considerations	12-14
Analyzing the FET Amplifier	12-15
12.6 Frequency Response of an Amplifier	12-17
High-Frequency Model of Common-Emitter Amplifier	12-17
Simplified hybrid-pi model	12-18
Figures of merit	12-19
High-Frequency Model of Common-Source Amplifier	12-20
Low-Frequency Response	12-21
12.7 Cascading Amplifier Stages	12-21
Coupling Cascaded Stages	12-21
Frequency Response of a Cascaded Amplifier	12-22
12.8 Power Amplifiers	12-22
Class A Power Amplifier	12-23
Collector Dissipation	12-23
Second-Harmonic Distortion	12-24
Push-Pull Power Amplifiers	12-24
Crossover Distortion	12-25
Heat Sinks	12-25
12.9 Feedback Amplifiers	12-26
Properties of Negative Feedback	12-26
Bandwidth	12-28
Distortion	12-28
Instability	12-28
Effects of Input and Output Resistance	12-28
12.10 Sinusoidal Oscillators	12-28
Phase-Shift Oscillators	12-29
Tuned-Circuit Oscillators	12-30
Frequency Stability	12-30
Crystal Oscillators	12-31

**13. OPERATIONAL AMPLIFIERS**

13.1	Introduction	13-1
13.2	The Basic OP AMP	13-1
	Ideal Characteristics	13-1
	Feedback	13-2
13.3	Applications	13-3
	Noninverting Amplifier	13-3
	Difference Amplifier	13-3
	Voltage Follower	13-4
	Summing Amplifier	13-4
	Logarithmic Amplifier	13-5
	Current-to-Voltage Converter	13-6
13.4	Practical Considerations	13-6
	Frequency Response and Compensation	13-6
	Offset	13-8
	Input offset current	13-8
	Input offset voltage	13-8
	Input offset current drift	13-8
	Input offset voltage drift	13-8
	Offset Correction	13-8
	Slewing Rate	13-8
13.5	OP AMP Circuitry	13-9
	Differential Amplifier	13-10
	Current Sink	13-11
	Cascading Stages	13-12
	DC Level Shifter	13-13
	Output Stage	13-12
13.6	Examples of Practical OP AMPS	13-14
13.7	Understanding the Data Sheet for an OP AMP	13-15

**14. DIGITAL CIRCUIT FUNDAMENTALS**

14.1	Introduction	14-1
14.2	Binary Numbers	14-1
	Decimal-to-Binary Conversion	14-2
	Decimal-to-Binary Fraction Conversion	14-2
	Binary-to-Decimal Conversion	14-3
	Binary-to-Decimal Fraction Conversion	14-3
14.3	Diode as a Switch	14-4
14.4	Transistor as a Switch	14-4
14.5	Basic Logic Circuits	14-6
	OR Gate	14-6
	Positive and negative logic	14-7
	OR gate circuit	14-7
	AND Gate	14-7
	AND gate circuit	14-8
	INVERTER	14-8
	INVERTER circuit	14-8
14.6	Half-Adder	14-10
14.7	INHIBIT and EXCLUSIVE OR Gates	14-11
14.8	NAND and NOR Gates	14-12
	NAND Gate	14-12
	NOR Gate	14-13



14.9	Multivibrators	14-14
	Bistable MV	14-15
	Monostable MV	14-16
	Astable MV	14-17
14.10	Dynamic Response of Diode and Transistor Switches	14-17
	Response of a Diode Switch	14-17
	Response of a Transistor Switch	14-18

## 15. DIGITAL INTEGRATED CIRCUITS

15.1	Introduction	15-1
15.2	DIC Terms and Parameters	15-1
	Fan In	15-1
	Fan Out	15-3
	Propagation Delay	15-3
	Speed	15-3
	Gate Dissipation	15-3
	Speed-Power Product	15-3
	Noise Margin	14-3
15.3	T <sup>2</sup> L	15-3
	Standard T <sup>2</sup> L	15-3
	Low-Power T <sup>2</sup> L	15-4
	High-Speed T <sup>2</sup> L	15-5
	Schottky-Clamped T <sup>2</sup> L	15-5
15.4	ECL	15-6
15.5	CMOS	15-7
	Inverter	15-7
	NOR Gate	15-7
	NAND Gate	15-8
15.6	Flipflops	15-8
	R-S Flipflop (Latch)	15-8
	Clocked R-S Flipflop	15-9
	J-K Flipflop	15-10
	T Flipflop	15-11
	D Flipflop	15-12
	Clear and Preset	15-12
	Master-Slave Flipflop	15-13
15.7	Counters	15-13
	Other Types of Counters	15-14
	Up counter	15-14
	Down counter	15-14
	Up-down counter	15-14
	Modulo counter	15-14
15.8	Shift Registers	15-15
	Serial register	15-15
	Serial-in, parallel-out register	15-15
	Parallel-in, serial-out register	15-15
	Parallel-in, parallel-out register	15-15
	Circulating register	15-16
15.9	Semiconductor Memories	15-17
	Types of Memories	15-18
	Access and Cycle Times	15-18
	ROM	15-18
	RAM	15-19

15.10	Digital-to-Analog and Analog-to-Digital Converters	15-19
	D/A Converter	15-20
	A/D Converter	15-23
15.11	The Microprocessor	15-23

## **16. POWER SUPPLIES**

16.1	Introduction	16-1
16.2	Rectifier Systems	16-1
	Half-Wave Rectifiers	16-2
	Filtering	16-3
	Full-Wave Rectifiers	16-4
	Tube Rectifiers	16-6
	Bridge Rectifiers	16-7
	Half-Wave Voltage Doublers	16-9
	Full-Wave Voltage Doublers	16-10
	Voltage Triplers and other Voltage Multipliers	16-10
	Load versus Load Resistance	16-12
	Combination Positive and Negative Supplies	16-12
16.3	Regulated Power Supplies	16-13
	Basic Types	16-15
	Series type	16-15
	Shunt type	16-15
	Switching type	16-15
16.4	Power-Supply Regulator Circuits	16-16

## **17. BATTERIES**

17.1	Introduction	17-1
	General Considerations	17-2
	Primary Battery	17-2
	Secondary Battery	17-3
	Reserve Batteries	17-3
17.2	Characteristics of Important Commercial Primary Batteries	17-5
	Basic Types	17-5
	Polarization of Dry Cells	17-5
	Performance Data	17-6
	American Standards Association Requirements	17-9
	Shelf Life	17-10
	Alkaline Batteries	17-11
	Mercury Dry Batteries	17-13
	Magnesium Dry Batteries	17-15
17.3	Characteristics of Important Commercial Secondary Types	17-17
	Lead-Acid Battery	17-17
	Nickel-Cadmium Batteries	17-19
	Silver-Zinc Batteries	17-22
17.4	Maintenance	17-24
	Lead-Acid Battery Maintenance	17-24
	Sealed Units	17-25
	Charging Nickel-Cadmium Batteries	17-26
	Charging Silver-Zinc Batteries	17-26
17.5	Special Types	17-27

Magnesium Batteries	17-28
Organic Depolarized Batteries	17-28
Cuprous Chloride-Magnesium Batteries	17-28
Silver Chloride-Magnesium Batteries	17-29
Air-Depolarized Batteries	17-29
Air-Zinc Batteries	17-30
17.6 Fuel Cells	17-31

## 18. MICROPROCESSORS AND MICROCOMPUTERS

18.1 Introduction	18-1
18.2 Transporting Digital Words	18-2
18.3 Introduction to Microprocessor Systems	18-4
Microcomputers	18-5
18.4 The 8085 Microprocessor	18-6
Registers	18-6
Stack Pointers	18-7
Program Counter	18-7
Arithmetic Logic Unit	18-8
Address Latch Incrementer-Decrementer	18-8
Data or Address Buffer	18-8
Instruction Register and Decoder	18-8
Timing and Control	18-8
Interrupt Control	18-8
Serial I/O Control	18-9
Basic System Timing	18-9
Multiplexing	18-10
Types of Addressing	18-11
18.5 The 8085 Instruction Set	18-13
Data Transfer Group	18-13
Arithmetic and Logic Group	18-16
Branch Control Group	18-18
Stack, I/O, and Machine Control Group	18-18
18.6 Software—Writing a Program	18-19
Steps in Writing a Program	18-19
Subroutines	18-21
18.7 Support Devices	18-22
Decoders	18-22
I/O Ports	18-22
Programmable Peripheral Interface (PPI)	18-22
Other Support Devices	18-22
18.8 Microprocessor Survey	18-23
Extended Addressing	18-23
Implied Addressing	18-24
Accumulator Addressing	18-24
Indexed Addressing	18-24
Relative Addressing	18-24
Comparison of Microprocessors	18-24
Intel 8080	18-24
Z80	18-24
6800	18-25
6502	18-25
18.9 Summary	18-26

**19. LOGIC ANALYZERS, LOGIC PROBES, AND SIGNATURE ANALYSIS**

19.1	Introduction to Logic Analyzers	19-1
19.2	Review of Digital Circuitry	19-1
19.3	Types of Faults in Digital Systems	19-3
	Hardware Faults	19-3
	No data	19-3
	Glitches	19-3
	Spikes	19-3
	Races	19-4
	Timing Errors	19-4
	Ringing	19-4
	Wrong level	19-5
	Software Faults	19-5
	Wrong instructions	19-5
	Latent faults	19-5
	Timing faults	19-5
	Memory defects	19-5
19.4	Requirements on Logic Analyzer	19-6
19.5	Basic Principles of the Logic Analyzer	19-6
	Memory	19-6
	Trigger Function	19-7
	Data Acquisition	19-7
	Display	19-8
19.6	Main Functions of a Logic Analyzer	19-10
	Data Acquisition	19-10
	The clock	19-10
	Data input	19-10
	Setup and hold times	19-11
	Clock qualifers	19-12
	Start function	19-12
	Trigger Function	19-12
	Manual triggering	19-13
	Internal triggering	19-13
	External triggering	19-15
	Trigger delay	19-15
	Display	19-17
	Synchronous and Asynchronous Sampling	19-20
19.7	Applications of the Logic Analyzer	19-21
	Exercises in Monitoring Data Transport via Bus Lines	19-21
19.8	Glossary of Logical Analyzer Terms	19-26
	Triggering	19-26
	Sampling	19-27
	Display	19-27
	Memory	19-27
	General	19-28
	Common Microprocessor Terminology	19-28
19.9	Logic Probe	19-30
19.10	Logic Pulser	19-30
19.11	Signature Analysis for Digital Troubleshooting	19-30
19.12	Basis of Signature Analysis	19-30
19.13	Test Routine in ROMs	19-31
	Testing Contents of Several ROMs	19-31
	Testing Unknown Boards	19-31



	Troubleshooting Techniques	19-32
19.14	The Design of a Signature Analyzer	19-32
	Review Questions and Answers about Signature Analysis	19-33

## 20. MICROWAVES

20.1	Basic Concepts	20-1
	Wavelength	20-1
	Microwave Region	20-1
	Power as a Fundamental Quantity	20-2
	Decibels	20-2
20.2	Transmission Lines	20-3
	Characteristic Impedance	20-3
	Coaxial Cable	20-4
	Parallel-Wire Transmission Line	20-5
	Stripline and Microstrip	20-5
	Other Shapes	20-5
	Waveguides	20-5
	Dielectric Transmission Lines	20-13
	Reflection Coefficient	20-13
	Voltage Standing Wave Ratio	20-14
	Smith Chart	20-15
	Power Transfer	20-19
	S Parameters	20-19
	Skin Effect	20-20
20.3	Circuit Components	20-21
	Passive Linear Reciprocal Components	20-22
	Terminations	20-22
	Attenuator	20-22
	Variable attenuator	20-22
	Phase shifter	20-24
	Coupler	20-24
	Hybrids	20-24
	Antennas	20-24
	Filters	20-24
	Passive Linear Nonreciprocal Components	20-25
	Isolator	20-25
	Limiter	20-28
	Active Linear Reciprocal Components	20-29
	Switch	20-29
	Attenuators	20-29
	Active Linear Nonreciprocal Components	20-29
	Amplifiers	20-29
20.4	Antennas	20-33
	Reciprocity	20-33
	Isotropic Radiator	20-33
	Antenna Patterns	20-33
	Polarization	20-34
	Directivity	20-34
	Gain	20-34
	Effective Area	20-34
	Beamwidth	20-35
	Estimating Beamwidth and Directivity	20-35

Antenna Arrays	20-36
Pattern Multiplication	20-37
Microwave Antennas	20-38

## 21. FIBER-OPTIC SYSTEMS

21.1 Introduction	21-1
21.2 Optical Fibers and Cables	21-4
Light Guiding	21-4
Attenuation	21-7
Fiber Bandwidth	21-12
Cabled Fibers	21-14
Fiber Connectoring	21-17
Splices	21-17
Connectors	21-18
Couplers	21-19
Connectoring mismatch losses	21-20
21.3 Optical Transmitters and Sources	21-21
Light-Emitting Diodes	21-23
Injection Laser Diodes	21-28
21.4 Optical Detectors and Receivers	21-32
Responsivity	21-34
Noise Equivalent Power	21-36
Sensitivity	21-38
21.5 Selected Bibliography	21-38

## 22. ACTIVE FILTERS

22.1 Introduction	22-1
22.2 Low-Pass and High-Pass Filters	22-2
First-Order Filters	22-2
Amplifying filters	22-3
Second-Order Filters	22-5
Effect of $Q$	22-5
Circuits	22-6
Design data	22-10
Circuit features	22-12
Third- and High-Order Filters	22-12
Single op amp third-order filter	22-12
Cascaded filters	22-13
Attenuation rate	22-14
22.3 Bandpass and Bandstop Filters	22-16
Broadband Filters	22-16
Narrowband Filters	22-17
Bandpass filter	22-17
Bandstop filter	22-20
22.4 References and Bibliography	22-22

## 23. DIGITAL MULTIMETERS

23.1 Basic Building Blocks	23-1
----------------------------	------

23.2	DC Input Amplifier/Attenuator	23-2
23.3	Internal Reference Voltage	23-3
23.4	AC Conversion	23-4
23.5	Measuring Resistance	23-6
23.6	A/D Conversion Techniques	23-7

## 24. OSCILLOSCOPE MEASUREMENTS

24.1	Introduction	24-1
24.2	Oscilloscope Controls	24-2
	Display System	24-2
	Beam finder	24-4
	Intensity	24-4
	Focus	24-4
	Trace rotation	24-4
	Vertical System	24-4
	Vertical position	24-5
	Input coupling	24-5
	VOLTS/DIV switch	24-5
	Variable volts per division	24-5
	Channel 2 inversion	24-5
	Vertical operating modes	24-5
	Horizontal System	24-6
	Horizontal position	24-6
	Seconds per division	24-6
	Variable seconds per division	24-7
	Horizontal magnification	24-7
	Horizontal operating mode	24-7
	Trigger System	24-7
	Trigger level and slope	24-8
	Variable trigger holdoff	24-8
	Trigger sources	24-9
	Trigger operating modes	24-10
	Trigger coupling	24-11
24.3	Oscilloscope Measurements	24-11
	Safety	24-11
	Getting Set Up	24-11
	Probes	24-12
	Probe compensation	24-12
	Probe handling	24-12
	Basic Waveforms	24-12
	Measurement Techniques	24-13
	Direct and derived measurements	24-14
	Period measurements	24-16
	Frequency measurements	24-16
	Rectangular-wave measurements	24-17
	Pulse measurements	24-17
	Phase measurements	24-18
	XY measurements	24-19
	Differential measurements	24-20
	Z axis measurement	24-20
	Delayed-sweep measurements	24-21
24.4	Effects of Instrument Performance	24-24

	Instrument Risetime and Measured Risetimes	24-24
	Relating Bandwidth and Risetime	24-25
	Measurement System Bandwidth	24-27
	Circuit Loading	24-27
24.5	Storage Oscilloscopes	24-28
	CRT Storage Oscilloscopes	24-28
	Bistable CRT storage	24-28
	Variable-persistence CRT storage	24-28
	Fast-transfer CRT storage	24-29
	Stored writing speed	24-29
	Digital Storage Oscilloscopes	24-30
	How digital storage works	24-30
	Comparing digital and CRT storage	24-31
	Digital storage oscilloscope specifications	24-32
24.6	Selecting an Oscilloscope	24-33
24.7	Bibliography	

**Index follows Chapter 24**

# List of Contributors

**JOSEPH R. CLEVELAND**

MCI

Richardson, Texas

**CHET HEYBERGER**

Tektronix, Inc.

Beaverton, Oregon

**KENNETH JESSEN**

Hewlett-Packard Co.

Loveland, Colorado

**GARY C. MEYER**

Philips Test and Measuring Instruments, Inc.

Minnetonka, Minnesota

**RICHARD J. PANKOW**

Norden Systems

Norwalk, Connecticut

**ROGER C. THIELKING**

Onondaga Community College

Syracuse, New York

**PAT ZAGARIA**

Philips Test and Measuring Instruments, Inc.

Mahwah, New Jersey



# Preface

This fully comprehensive sourcebook—a condensation of the authors' successful *Handbook for Electronics Engineering Technicians*, Second Edition—is designed to meet the day-to-day needs of electronics technicians. While useful to engineers as well, it does not require an extensive background in high-level engineering principles and techniques. All that is required is a background no more extensive than an education in an electronics technical school or a two-year community college. And in many cases, because of the sourcebook's simplified approach, even a lesser background would be sufficient.

The sourcebook treats fundamental topics in discrete circuits, and also in analog and digital integrated circuits, from the point of view of practical applications. Each topic is illustrated with practical, numerical, worked-out examples that can be applied to the reader's own particular problems. Numerous practical examples illustrate, for example, how to choose the proper resistor, capacitor, transistor, integrated circuit, or operational amplifier and how to find current in, or voltage across, an element in dc and ac circuits.

Each one of the twenty-four in-depth sections follows the same practical, concise format that has been developed to help readers find the information they need quickly and easily. The general approach to each topic in a section is as follows:

1. Definition of terms and parameters
2. Breakdown of the types of characteristics of components
3. Analysis of the basic and special functions
4. Detailed practical problems and clearly worked-out solutions
5. Clarifying charts, tables, nomographs, and illustrations

With this easy-to-follow format, every effort has been made to ensure that this sourcebook will have the greatest usefulness to its readers.

All mathematics, both in theory sections and in worked-out problems, has been kept to the level of relatively simple algebra or arithmetic. No calculus is employed.

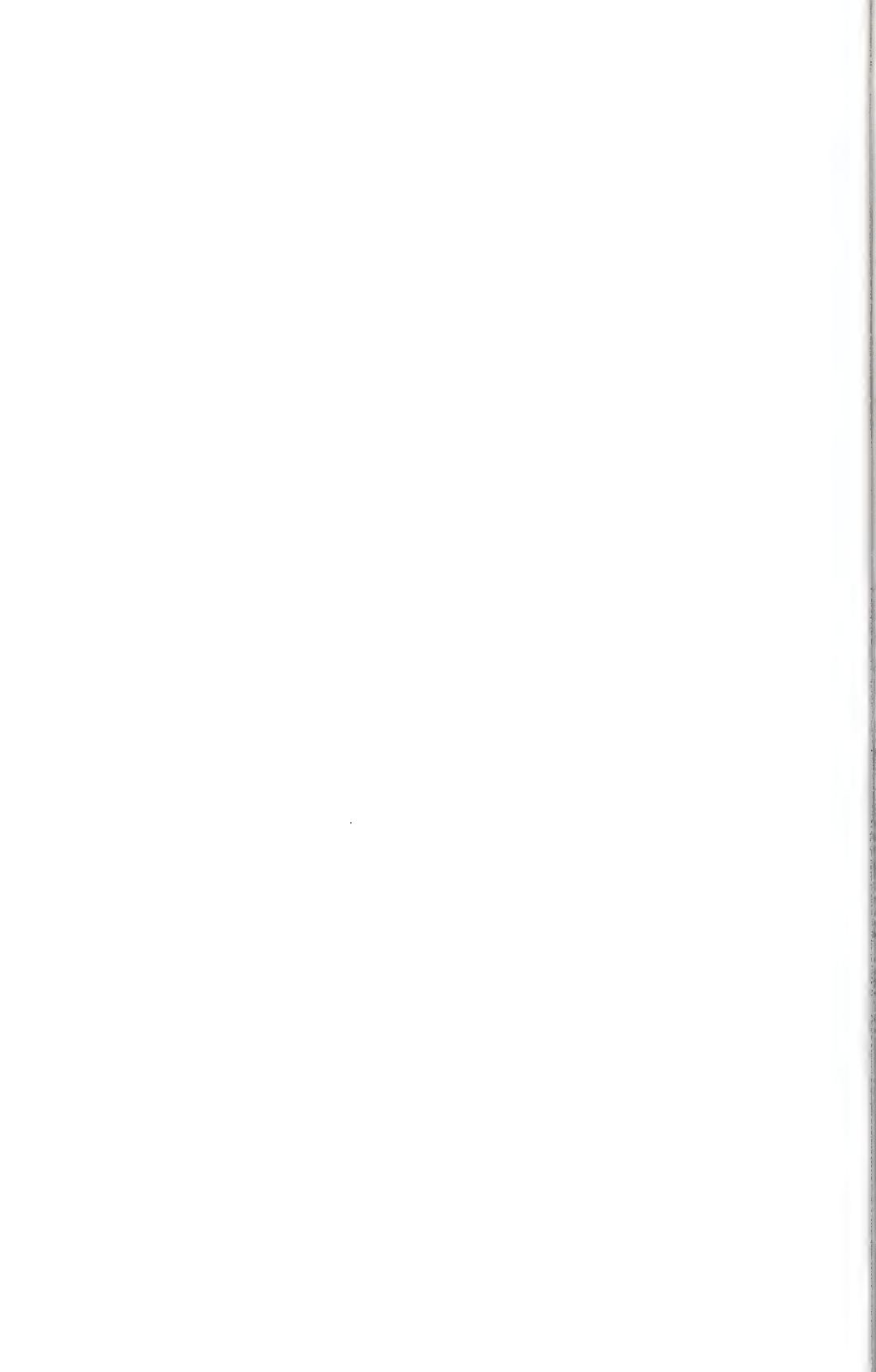
The following chapters were written by Professor Seidman: Chapters 1, 2, 6, 8, 9, and 12 to 15.

The following chapters have been written or edited by Mr. Kaufman: Chapters 3 to 5, 7, 10, 11, 16, and 17.

The editors wish to gratefully acknowledge the cooperation and assistance of various individuals and companies in providing manuscript material, photographs, tables, graphs, and other technical data. Due credit is given within the book as applicable. We wish particularly to thank the following persons for their invaluable contributions: Donald R. Phillips of RCL Electronics, Inc.; Dr. Howard J. Strauss of Gould, Inc.; J. A. "Sam" Wilson of Bank Wilson Services; Joshua A. Hauser of Lambda Electronics, Inc.; Raymond C. Miles; Bull Stutz; and the late Louis Toth. In addition, we wish to thank Sharon Wilson for her considerable efforts in typing and editorial services.

Milton Kaufman  
Arthur H. Seidman





# Chapter 1

## Characteristics of Resistors

### 1.1 INTRODUCTION

Every physical material impedes the flow of electric current to some degree. Materials such as copper offer hardly any resistance to current flow; copper, therefore, is called a *conductor*, or a material having negligible resistance. Other materials, such as ceramic, which offer extremely high resistance to current flow are referred to as *insulators*.

In electric and electronic circuits, there is a need for materials with specific values of resistance in the range between that of a conductor and an insulator. These materials are called *resistors* and their values of resistance are expressed in ohms (represented by the Greek letter omega,  $\Omega$ ). The large variety of types and forms of discrete resistors available are indicated in Fig. 1.1.

Resistors may be classified as being *fixed* or *variable* in their value. Variable resistors are commonly referred to as *potentiometers*, or *pots*. Electrical symbols for fixed resistors and pots are illustrated in Fig. 1.2.

Resistors may also be classified as *linear* and *nonlinear*. For a linear resistor, as the voltage across it varies, the current flowing varies by a proportionate amount. The behavior of a nonlinear resistor is such that as the voltage varies, the current change is not proportional to the voltage change. Resistors generally used in circuits are linear. For special applications, to be described later, nonlinear resistors are available.

### 1.2 GENERAL DESCRIPTION

The resistance of any material is given by the following expression:

$$R = \frac{\rho L}{A} \quad (1.1)$$

where  $R$  = resistance,  $\Omega$

$\rho$  = resistivity of the material,  $\Omega\text{-cm}$

$L$  = length of material, cm

$A$  = cross-sectional area of material,  $\text{cm}^2$

Resistivity  $\rho$  (Greek letter rho) is an inherent property of materials. Values of  $\rho$  for some commonly used materials are summarized in Table 1.1.

Equation (1.1) illustrates two important facts. For a material of a given resistivity, the resistance varies *directly* with length  $L$  and *inversely* with cross-sectional area  $A$ . For example, a wire of long length has greater resistance than a short-length wire. Also, a wire of large diameter (large cross-sectional area) has less resistance than a wire with a small diameter.

## 1-2 Characteristics of Resistors



**Fig. 1.1** A sampling of some discrete resistors available to the user. (Courtesy Victoreen Instrument Division)

The voltage and current in a resistor are related by Ohm's law:

$$I = \frac{E}{R} \quad (1.2a)$$

$$E = IR \quad (1.2b)$$

$$R = \frac{E}{I} \quad (1.2c)$$

where  $E$  = voltage across resistor, V

$I$  = current flowing in resistor, A

Power  $P$  (in watts, W) dissipated in a resistor may be expressed by any of the following expressions:

$$P = EI \quad (1.3a)$$

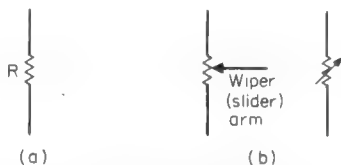
$$= I^2 R \quad (1.3b)$$

$$= \frac{E^2}{R} \quad (1.3c)$$

The application of the preceding equations is illustrated in the following examples.

**example 1.1** Calculate the resistance of a wire whose length is 10 m and cross-section area is  $0.1 \text{ cm}^2$  if the material of the wire is (a) copper and (b) Nichrome.

**solution** From Table 1.1, for copper  $\rho = 1.7 \times 10^{-6} \Omega\text{-cm}$  and for Nichrome,  $\rho = 100 \times 10^{-6}$ . Because  $1 \text{ m} = 100 \text{ cm}$  (units in an equation must always be consistent),  $L = 10 \times 100 \text{ cm} = 1000$



**Fig. 1.2** Electrical symbols for resistors: (a) Fixed. (b) Variable (pot).

**TABLE 1.1 Resistivities of Some Commonly Used Materials**

Material	$\rho$ , $\Omega\text{-cm}$
Silver	$1.5 \times 10^{-6}$
Copper	$1.7 \times 10^{-6}$
Aluminum	$2.6 \times 10^{-6}$
Carbon (graphite)	$30 \times 10^{-6}$ to $190 \times 10^{-6}$
Nichrome	$100 \times 10^{-6}$
Glass	$10^{10}\text{--}10^{14}$

cm. Substitution of the given values in Eq. (1.1),  $R = \rho L/A$ , yields

$$(a) R_{\text{copper}} = 1.7 \times 10^{-6} \Omega \text{ cm} \times (1000 \text{ cm}) / (0.1 \text{ cm}^2)$$

$$= 1.7 \times 10^{-2} \Omega$$

$$(b) R_{\text{Nichrome}} = 100 \times 10^{-6} \times 1000 / 0.1 = 1 \Omega$$

**example 1.2** If the current flowing in the wires of Example 1.1 is 3 A, calculate the dissipated power in each wire.

**solution** Using Eq. (1.3b)  $P = I^2 R$ ,  $P_{\text{copper}} = 3^2 \times 1.76 \times 10^{-2} = 9 \times 1.76 \times 10^{-2} = 0.158 \text{ W}$ .  
 $P_{\text{Nichrome}} = 3^2 \times 1 = 9 \times 1 = 9 \text{ W}$ .

### 1.3 RESISTOR TERMS AND PARAMETERS

In this section, commonly used terms and parameters for characterizing fixed and variable resistors are defined. Where appropriate, typical curves showing the variation in resistance with temperature and other quantities of interest are supplied.

**RESISTANCE** The unit of resistance is the ohm ( $\Omega$ ). Generally, resistance values in thousands of ohms is expressed in kilohms ( $k\Omega$ ) and millions of ohms in megohms ( $M\Omega$ ). Nominal values of resistors are generally based on  $25^\circ\text{C}$  (room temperature) operation.

**TOLERANCE** Tolerance expresses the maximum deviation in resistance from its nominal value. For example, if the tolerance of a  $1000\text{-}\Omega$  resistor is  $\pm 10$  percent, this denotes that the actual value of resistance is in the range of  $1000 - 0.1 \times 1000 = 900 \Omega$  to  $1000 + 0.1 \times 1000 = 1100 \Omega$ .

**TEMPERATURE COEFFICIENT OF RESISTANCE** The temperature coefficient of resistance (abbreviated TCR, or *Tempco*) indicates how resistance changes with temperature. It is expressed as a percentage change in the nominal value at  $25^\circ\text{C}$  for each degree Celsius or in parts per million per degree Celsius ( $\text{ppm}/^\circ\text{C}$ ) of the resistor. The TCR may be positive or negative. Typical curves showing the variation in resistance with temperature are illustrated in Fig. 1.3. In this figure, a *logarithmic (log) scale* is used for plotting nominal resistance (ohms). A log scale permits a large range of values (in this case,  $10^1$  to  $10^8 \Omega$ ) to be compressed in a convenient scale.

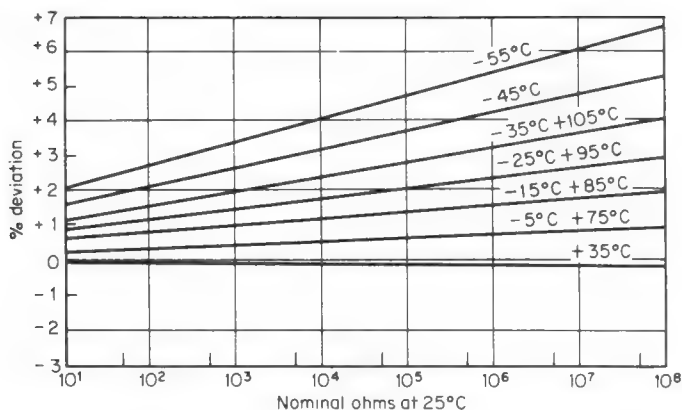
**example 1.3** A  $1000\text{-}\Omega$  resistor has a TCR of  $+1000 \text{ ppm}/^\circ\text{C}$ . Calculate its resistance at  $125^\circ\text{C}$ .

**solution** The quantity  $1000 \text{ ppm}$  is equal to  $10^3/10^6 = 0.001$ . The change in resistance from room temperature ( $25^\circ\text{C}$ ) is equal to the product of the difference in temperature ( $125 - 25^\circ$ ), the TCR, and the nominal value of resistance at  $25^\circ\text{C}$  ( $1000 \Omega$ ). Hence, the change in resistance =  $(125 - 25) \times 0.001 \times 1000 = 100 \Omega$ . The resistance value at  $125^\circ\text{C}$  is, therefore,  $1000 + 100 = 1100 \Omega$ .

**POWER RATING** Power rating is the maximum continuous power, in watts, that a resistor can dissipate at a temperature as high as  $70^\circ\text{C}$ . At temperatures beyond  $70^\circ\text{C}$ , the power rating of the resistor is reduced, or *derated*. A typical derating curve is given in Fig. 1.4.

**example 1.4** A particular resistor is rated at  $60 \text{ W}$ . Using the derating curve of Fig. 1.4, determine the power rating of the resistor when operating at  $130^\circ\text{C}$ .

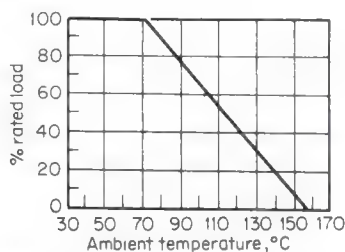
## 1-4 Characteristics of Resistors



**Fig. 1.3** Typical curves showing the variation in nominal resistance with temperature. (Courtesy Allen-Bradley)

**solution** From Fig. 1.4, at 130°C, the rated load is 30 percent. Expressing 30 percent by the decimal 0.3, the rating of the resistor at 130°C is  $0.3 \times 60 = 18 \text{ W}$ .

In specifying the power rating of a resistor, for conservative operation it is usual practice to increase the actual dissipated power by a factor of 2. For example, if the calculated dissipation is 0.5 W, specify a  $0.5 \times 2 = 1\text{-W}$  resistor.



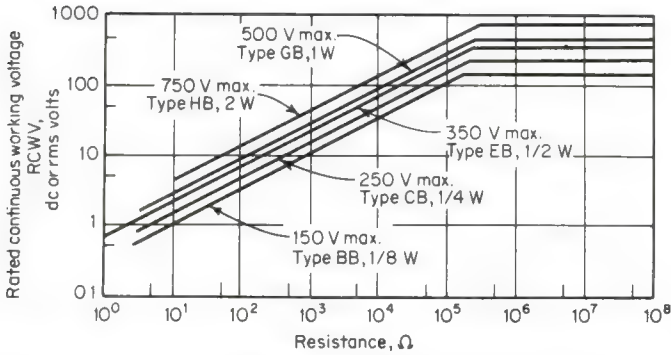
**Fig. 1.4** Typical derating curve for a resistor. Ambient temperature refers to the temperature of the environment in which the resistor is operating.

**RATED CONTINUOUS WORKING VOLTAGE** The rated continuous working voltage (RCWV) is the maximum voltage that can be safely applied to a resistor. Typical curves for RCWV versus resistance are shown in Fig. 1.5. For example, the maximum rated continuous working voltage for a  $50\text{-}\Omega$  2-W resistor is 10 V.

**CRITICAL RESISTANCE** The critical resistance  $R_c$  is the value of resistance where the maximum voltage and power rating occur simultaneously. For example, for a resistor rated at 500 V and 1 W, from Eq. (1.3c),  $R_c = E^2/P = 500^2/1 = 250\,000\ \Omega$ .

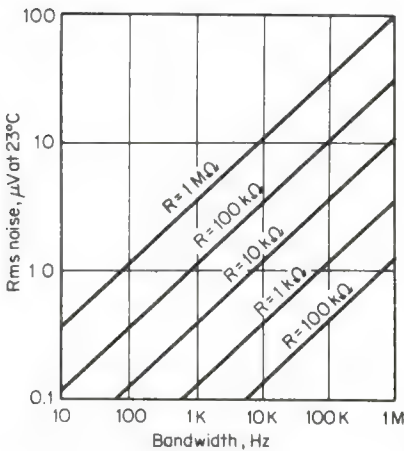
**NOISE** Because of the molecular structure of matter, electrons in a material exhibit a random motion. As a result, random voltages are produced which are referred to as *noise*. The noise increases with increasing resistance values, operating temperature, and the bandwidth of the circuit to which the resistor is connected. Typical curves of noise voltage (microvolts) at 25°C versus bandwidth (hertz) are illustrated in Fig. 1.6. It is seen that as the resistance value or bandwidth is increased, the noise voltage becomes greater.

**FREQUENCY EFFECTS** A *model* (equivalent circuit) of a resistor operating at high frequencies is illustrated in Fig. 1.7. In series with the desired resistance  $R$  is induct-

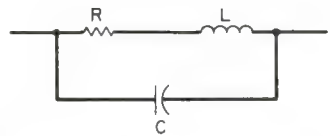


**Fig. 1.5** Typical curves of RCWV for different power ratings as a function of resistance. (Courtesy Allen-Bradley)

ance  $L$ ; shunted across  $R$  and  $L$  is capacitance  $C$ . Inductance  $L$  arises from the type of construction and connecting leads of the resistor. Capacitance  $C$  is present also because of the resistor's construction and the capacitance of the connecting leads. These undesired elements,  $L$  and  $C$ , are referred to as *parasitics*.



**Fig. 1.6** Typical curves of noise for different values of resistance as a function of bandwidth. (Courtesy MEPCO/ELECTRA, Inc.)



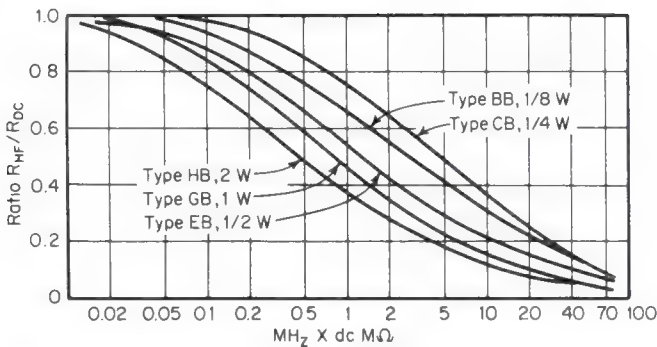
**Fig. 1.7** A model of a resistor operating at high frequencies.

Owing to the parasitics in a resistor, the useful frequency range is limited. Typical high-frequency characteristics of a resistor are illustrated in Fig. 1.8. Expression  $R_{h-f}/R_{dc}$  is the ratio of the resistance at high frequencies,  $R_{h-f}$  to the resistance at direct current and low frequencies,  $R_{dc}$ . It is seen that as frequency is increased,  $R_{h-f}/R_{dc}$  is reduced.

**DRIFT** Drift, or *time stability*, refers to the change in resistance value over a time interval of use, such as 1000 h. High-value resistors exhibit more drift than low-value units. Short-term drift is generally more pronounced than long-time drift.

**TAPER** The taper of a pot refers to the variation of its resistance as a function of the rotation of the wiper, or slider, arm. Examples of taper curves are illustrated in Fig. 1.9.

## 1-6 Characteristics of Resistors



**Fig. 1.8** Typical high-frequency characteristics of different-type fixed resistors. (Courtesy Allen-Bradley)

A *linear taper* is exhibited by curve *a*. For this taper, the variation in resistance is directly proportional to the displacement of the wiper. *Nonlinear tapers* are indicated by curves *b* and *c*. Curve *b* represents a logarithmic taper, and curve *c* a special nonlinear taper.

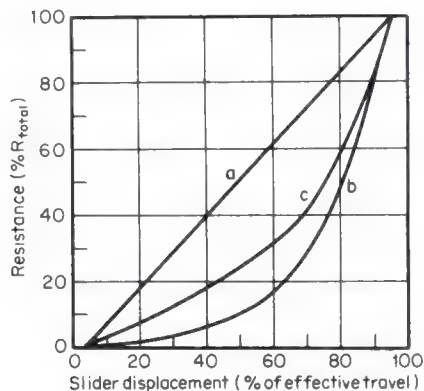
**RESOLUTION** The resolution of a pot is the smallest change in resistance that can be realized as the wiper arm is rotated.

**END RESISTANCE** End resistance is the resistance between the wiper and end terminals with the wiper positioned at the corresponding end point.

**CONTACT RESISTANCE** Contact resistance is the resistance between the wiper terminal and the resistance element in contact with the wiper.

**WIPER (SLIDER) CURRENT** Wiper current is the maximum current that may flow in or out of the wiper terminal.

**SETTING STABILITY** This parameter indicates the *repeatability* of a pot setting to a given resistance value.



**Fig. 1.9** Examples of taper curves for potentiometers. (Courtesy MEPCO/ELECTRA, Inc.)



TABLE 1.2 Summary of Fixed and Variable Resistor Types

Category	Type	Key property	Power	Temperature coefficient, °C	Resistance range
General purpose ≥ 5% tolerance ≥ 200 ppm	Carbon composition	Cost	$\frac{1}{8}$ -2 W	> 500	1 $\Omega$ - 100 M $\Omega$
	Molded wire-wound	Tempco	$\frac{1}{2}$ -2 W	≥ 200	0.1 $\Omega$ -2.4 k $\Omega$
	Ceramic wire-wound	Low voltage W	2-50 W	≥ 200	0.1 $\Omega$ -30 k $\Omega$
	Metal glaze	Flexibility	$\frac{1}{8}$ -5 W	200	4.3 $\Omega$ -1.5 M $\Omega$
	Tin oxide	Reliability	$\frac{1}{8}$ -20 W	200	4.3 $\Omega$ -2.5 M $\Omega$
Semiprecision > 1 < 5% ≤ 200 ppm	Carbon film (import)	Cost	$\frac{1}{4}$ -2 W	> 200	10 $\Omega$ - 1 M $\Omega$
	Cermet film	Stability	$\frac{1}{4}$ -3 W	150	10 $\Omega$ -10 M $\Omega$
	Metal glaze Tin oxide	Flexibility Stability	$\frac{1}{8}$ -2 W $\frac{1}{8}$ -2 W	≤ 200 ≤ 200	1 $\Omega$ -1.5 M $\Omega$ 4.3 $\Omega$ -1.5 M $\Omega$
Power ≥ 2 W	Ceramic wire-wound	Cost	2-50 W	≥ 200	0.1 $\Omega$ -30 k $\Omega$
	Axial lead coated WW Tubular and flat WW	Auto insertion	$\frac{1}{2}$ -15 W 4-250 W	≤ 50 ≤ 100	0.1 $\Omega$ -175 k $\Omega$ 0.1 $\Omega$ - 1 m $\Omega$
Precision ≤ 1% ≤ 100 ppm	Metal film	Tolerance	$\frac{1}{10}$ -1 W	20	0.1 $\Omega$ -1 M $\Omega$
	Metal glaze	Environment	$\frac{1}{10}$ -1 W	≤ 100	1 $\Omega$ -1 M $\Omega$
	Tin oxide	Power	$\frac{1}{10}$ - $\frac{1}{10}$ W	≤ 100	10 $\Omega$ -1 M $\Omega$
	Thin film	Size, networks	$\frac{1}{10}$ -5 W	≤ 100	10 $\Omega$ -100 M $\Omega$
	Encaps. wire-wound	Power, Tempco	$\frac{1}{20}$ -1 W	≤ 20	0.1 $\Omega$ - 1 M $\Omega$
Ultraprecision ≤ 0.5% ≤ 25 ppm	Thin film	Flexibility	$\frac{1}{20}$ - $\frac{1}{2}$ W	≤ 25	20 $\Omega$ -1 M $\Omega$
	Encaps. wire-wound	Noise	$\frac{1}{20}$ -1 W	≤ 20	0.1 $\Omega$ - 1 M $\Omega$
Variable devices (pots, trimmers)	Wire-wound	Tempco	5 at 70°C	± 20	10 $\Omega$ -100 k $\Omega$
	Conductive plastic	Rotational life	2 at 70°C	± 250-500	1 k $\Omega$ -100 k $\Omega$
	Cermet	Environmental	12 at 70°C	± 250-500	500 $\Omega$ -2 M $\Omega$
	Carbon	Cost	5 at 70°C	± 300-2000	100 $\Omega$ -2 M $\Omega$
Networks	Thick film	Cost	≤ 2 W/pkg.	≤ 200	10 $\Omega$ -10 M $\Omega$
	Thin film	Performance	≤ 2 W/pkg.	≤ 100	10 $\Omega$ -1 M $\Omega$

Reprinted from *Electronics*, May 30, 1974; copyright McGraw-Hill, Inc., 1974.

## 1.4 FIXED RESISTORS

Fixed resistors are manufactured in four basic types:

1. Carbon composition
2. Metal film
3. Carbon film
4. Wire-wound

Resistors can also be classified in terms of their tolerance:

1. General purpose: tolerance 5 percent or greater
2. Semiprecision: tolerance between 1 and 5 percent
3. Precision: tolerance between 0.5 and 1 percent
4. Ultraprecision: tolerance better than 0.5 percent

A summary of different resistor types and their key properties appears in Table 1.2.

**CARBON COMPOSITION** The carbon-composition resistor is perhaps the most widely used fixed resistor in discrete circuits. Composition resistors are available in resistance values from  $1\ \Omega$  to  $100\ \text{M}\Omega$  and typical power ratings of  $\frac{1}{8}$  to 2 W. Their temperature coefficient is high (greater than  $500\ \text{ppm}/^\circ\text{C}$ ), and their cost is low.

A cutaway view of a carbon resistor is shown in Fig. 1.10. The resistance material is a form of carbon, such as graphite, embedded in a binder. The resistance and in-

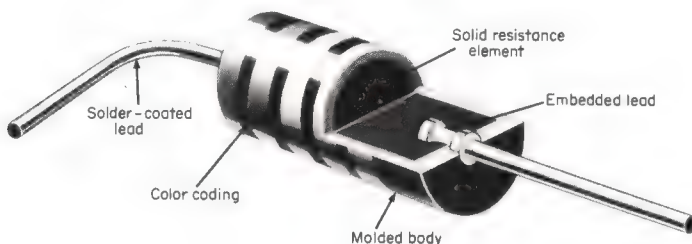


Fig. 1.10 Cutaway view of a carbon-composition resistor. (Courtesy Allen-Bradley)

sulation material, and the wire leads, are molded simultaneously under high temperature and pressure. The result is referred to as a slug-type structure. Different resistance values are obtained by varying the carbon and filler content.

**METAL FILM** Metal-film resistors are available as thin- and thick-film-type components. In the thick-film category are the tin-oxide, metal-glaze, cermet, and bulk-film resistors. Each type will be considered in this section.

**Thin film** The resistance element in a thin-film resistor is a film having a thickness in the order of one-millionth of an inch. (A thick film has a thickness greater than one-millionth of an inch.) Typically, the thin film is deposited on a ceramic substrate under a high vacuum; the technique is referred to as *vacuum deposition* (see Chap. 9). Metals used for deposition include nickel and chromium. Some characteristics of the thin-film resistor are: resistance range of  $10\ \Omega$  to  $1\ \text{M}\Omega$ ; tolerance better than 0.5 percent, TCR less than  $25\ \text{ppm}/^\circ\text{C}$ , power rating up to 5 W; low noise.

**Tin oxide** Tin oxide, in vapor form, is usually deposited on a ceramic substrate under high temperature. The vapor reacting with the substrate, which is heated, results in a tightly formed resistance film. Some characteristics of the tin-oxide resistor are: resistance range of a few ohms to  $2.5\ \text{M}\Omega$ , tolerance better than 1 percent, TCR less than  $200\ \text{ppm}/^\circ\text{C}$ , power rating up to 2 W, good stability.

**Metal glaze** A powdered glass and fine metal particle (palladium and silver) mixture is deposited on a ceramic substrate. The combination is then fired at a high temperature (typically  $800^\circ\text{C}$ ). This results in a fusion of metal particles to the substrate.

Some characteristics of the metal-glaze resistor are resistance range of a few ohms to 1.5 M $\Omega$ , tolerance better than 1 percent, TCR as low as 20 ppm/ $^{\circ}\text{C}$ , power rating up to 5 W, good stability.

**Cermet** A cermet-film resistor is made by screening (see Chap. 9) a mixture of precious metals and binder material on a ceramic substrate. Similar to the manufacture of a metal-glaze resistor, the combination is then fired at a high temperature. An example of a cermet film resistor is illustrated in Fig. 1.11. Some characteristics of this type of resistor are: resistance range of 10  $\Omega$  to 10 M $\Omega$ , tolerance as low as 1 percent, TCR in the order of 100 ppm/ $^{\circ}\text{C}$ , power rating up to 3 W, good stability.

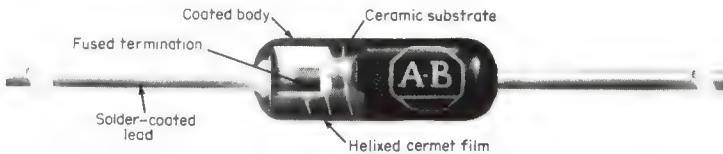


Fig. 1.11 Cutaway view of a cermet-film resistor. (Courtesy Allen-Bradley)

**Bulk film** In this resistor, the metal film is etched on a glass substrate (Fig. 1.12). Because of their unequal coefficients of expansion, the metal film is compressed slightly by the glass substrate. The compressed film has a negative temperature coefficient, which cancels out the inherent positive temperature coefficient of the film. As a result, the bulk-film resistor has a TCR close to zero. Some characteristics of the bulk-film resistor are resistance range of 30  $\Omega$  to 600 k $\Omega$ , tolerance as low as 0.005 percent, TCR in the order of 1 ppm/ $^{\circ}\text{C}$ , power rating up to 1 W, very low noise, good stability.

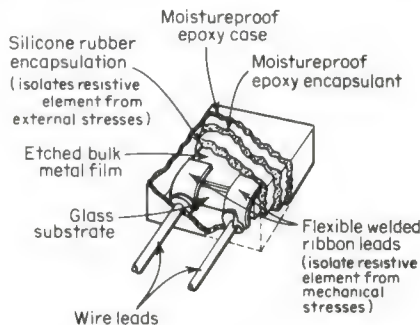


Fig. 1.12 Cutaway view of a bulk-film resistor. (Courtesy Vishay Resistor Products)

**CARBON FILM** This resistor is manufactured by depositing a carbon film on a ceramic substrate (Fig. 1.13). Some characteristics of carbon-film resistors are resistance range of 10  $\Omega$  to 10 M $\Omega$ , tolerance 5 percent or greater, TCR in the order of 150 ppm/ $^{\circ}\text{C}$ , power rating up to 2 W, generally less noisy than the carbon-composition resistor, low cost.

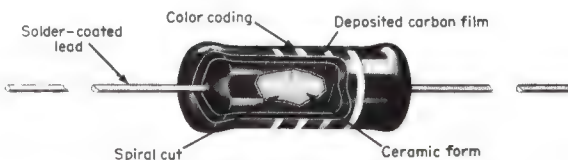
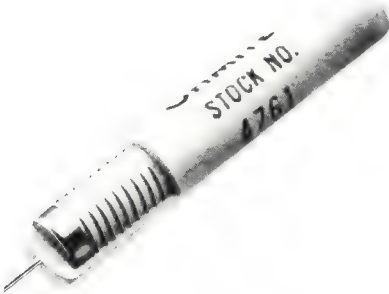
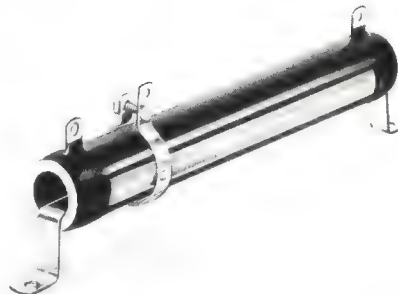


Fig. 1.13 Cutaway view of a carbon-film resistor. (Courtesy Piher International Corporation)



**Fig. 1.14** Cutaway view of a power-style molded vitreous enamel wire-wound resistor. (Courtesy Ohmite Manufacturing Company)



**Fig. 1.15** Adjustable power-style wire-wound resistor. (Courtesy Ohmite Manufacturing Company)

**WIRE-WOUND** The wire-wound resistor enjoys a broad spectrum of applications. Made in many shapes and sizes, it is used as an ultraprecision resistor in instrumentation and as a power resistor in industrial applications. Examples of its construction are illustrated in Figs. 1.14 to 1.16.

The *power-style* wire-wound resistor is made by winding a single-layer length of special alloy wire, in the form of a coil, around an insulating core. The unit is then covered with a coating, such as vitreous enamel (an inorganic glasslike mixture) or silicone. The coating protects the winding against moisture and breakage.

The resistance wire used must have carefully controlled resistance per unit length of wire and low-temperature coefficient, and be able to operate at high temperatures. Alloys used include nickel-chromium-aluminum (800 alloy) and nickel-chromium-iron (Nichrome). The cylindrical core is ceramic, steatite, or a vitreous material. In the *precision-style* wire-wound resistor, typically a multilayer coil is wound on an epoxy form, or *bobbin*.

Because the wire-wound resistor acts like a coil, its inductance (as well as the capacitance between coil turns) is a problem at high-frequency operation. A number of techniques are utilized to minimize inductance. In one method, one-half the resistance wire is wound in one direction and the other half in the opposite direction. This type of winding is referred to as *bifilar*.

In another method, used in precision-style resistors, a thick-film *serpentine pattern* is deposited on a ceramic core (Fig. 1.17). Similar in behavior to the bifilar winding,



**Fig. 1.16** Examples of precision-style wire-wound resistors. (Courtesy Shallicross)



**Fig. 1.17** A serpentine resistor pattern used to reduce inductance. (Courtesy Caddock Electronics)

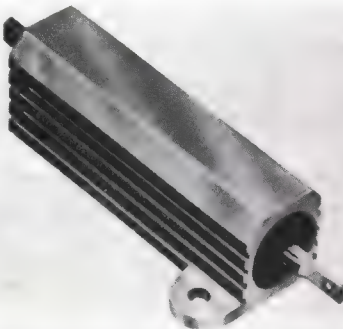
opposite magnetic fields produced by current flowing in adjacent resistance paths cancel each other. The result is a resistor with virtually no inductance.

Precision-style wire-wound resistors are available having the following characteristics: resistance range of a fraction of an ohm to 10 M $\Omega$ , tolerance better than 0.5 percent, TCR less than 20 ppm/ $^{\circ}\text{C}$ , power rating in the order of 2 W.


Some characteristics of power-style wire-wound resistors are: resistance range of less than an ohm to greater than a megohm; tolerance in the range of  $\pm 5$  to  $\pm 20$  percent; TCR as low as 5 ppm/ $^{\circ}\text{C}$ ; power rating as high as 1500 W. To increase its power rating, the resistor is sometimes placed in a metal housing, such as aluminum, illustrated in Fig. 1.18. For some types of resistors, the power rating of a resistor in a metal housing is increased by a factor of 2. For the realization of the increased power rating, however, the housing must be mounted on a metal chassis or suitable heat sink.

## 1.5 COLOR CODING OF RESISTORS

Figure 1.19 illustrates the standard color code for composition and some axial-type resistors. Different color bands are used to designate the resistance value and toler-



**Fig. 1.18** A power resistor mounted in an aluminum housing. (Courtesy RCL Electronics, Inc.)



Color	Digit	Multiplier	Tolerance
Black	0	1	—
Brown	1	10	$\pm 1\%$
Red	2	100	$\pm 2\%$
Orange	3	1000	—
Yellow	4	10 000	—
Green	5	100 000	—
Blue	6	1 000 000	—
Violet	7	10 000 000	—
Gray	8	—	—
White	9	—	—
Gold	—	0.1	$\pm 5\%$
Silver	—	—	$\pm 10\%$
No color	—	—	$\pm 20\%$

**Fig. 1.19** Standard color code for composition and some axial-type resistors.

ance. The first two bands denote the first and second digits of the resistance value and the third band indicates how many zeros follow the first two digits. Tolerance is given by the fourth band. For example, a resistor with the following colored bands, yellow-violet-orange-silver, denotes a 47 000- $\Omega \pm 10$  percent tolerance resistor.

For film and wire-wound resistors, in general, the resistance value and tolerance are stamped on the body of the resistor. Occasionally, a manufacturer may use his own code. For this reason, it is good practice to consult the manufacturer's catalog or data sheet.

## 1.6 CONNECTING RESISTORS

In this section we examine how fixed resistors are connected in series and parallel.

**RESISTORS IN SERIES** Figure 1.20 shows  $n$  resistors connected in series. In a series circuit, the current  $I$  flowing in each resistor is the same. The total equivalent resistance of the series circuit  $R_{TS}$  is equal to the sum of the individual resistors:

$$R_{TS} = R_1 + R_2 + \cdots + R_n \quad (1.4)$$

## 1-12 Characteristics of Resistors

If  $n$  equal resistors are connected in series, the total equivalent resistance is equal to the product of the value of an individual resistor and  $n$ :

$$R_{TS} = nR \quad (1.5)$$

The maximum voltage  $E_{\max}$  that can be applied across a series circuit of resistors is equal to the sum of the rated continuous working voltage RCWV of each resistor:

$$E_{\max} = (\text{RCWV})_1 + (\text{RCWV})_2 + \cdots + (\text{RCWV})_n \quad (1.6)$$

**example 1.5** Three resistors of 10, 20, and 30  $\Omega$  are connected in series. The current flowing in the circuit is 2 A, and the RCWV for each resistor is 500 V. Determine (a) the total equivalent resistance of the circuit, (b) the power dissipated in each resistor, (c) the maximum voltage that can be impressed across the series circuit. If the three resistors are replaced by a single equivalent resistor, determine (d) its power and maximum voltage ratings.

**solution** (a) By Eq. (1.4)  $R_{TS} = 10 + 20 + 30 = 60 \Omega$ .

(b) By Eq. (1.3b),  $P = I^2 R$ , we have

$$P_{10} = 2^2 \times 10 = 4 \times 10 = 40 \text{ W}$$

$$P_{20} = 2^2 \times 20 = 4 \times 20 = 80 \text{ W}$$

$$P_{30} = 2^2 \times 30 = 4 \times 30 = 120 \text{ W}$$

(c) By Eq. (1.6),  $E_{\max} = 500 + 500 + 500 = 1500 \text{ V}$ .

(d) In part a we found that the equivalent resistance is 60  $\Omega$ . The power rating, therefore, is  $2^2 \times 60 = 4 \times 60 = 240 \text{ W}$ . (Note: This value must equal the sum of the dissipated power in each individual resistor found in part b.) From part c, the maximum voltage rating is 1500 V.

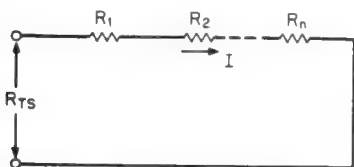


Fig. 1.20 Resistors connected in series.

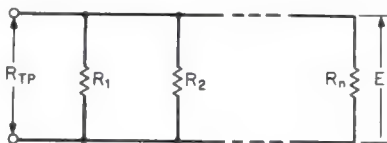


Fig. 1.21 Resistors connected in parallel (shunt).

**example 1.6** A 90- and a 10- $\Omega$  resistor are connected in series. The impressed voltage across the series combination is 1000 V. Determine the voltage across each resistor.

**solution** By voltage division, the voltage across each resistor is proportional to its resistance value. For the 10- $\Omega$  resistor,  $E_{10} = 1000 \times 10/(10 + 90) = 100 \text{ V}$ ; for the 90- $\Omega$  resistor,  $E_{90} = 1000 \times 90/(10 + 90) = 900 \text{ V}$ .

What is important in this example is that in a series circuit the voltage across the larger resistor (90  $\Omega$ ) is much greater than that across the smaller (10  $\Omega$ ) resistor. The RCWV of the 90- $\Omega$  resistor, therefore, must be at least 900 V. For the 10- $\Omega$  resistor, an RCWV of 100 V is adequate.

**RESISTORS IN PARALLEL** Figure 1.21 shows  $n$  resistors connected in parallel. In a parallel circuit, the voltage  $E$  across each resistor is the same. It is convenient to consider two resistors at a time when calculating the equivalent resistance  $R_{TP}$  of a parallel circuit. The equivalent resistance  $R$  of two resistors  $R_1$  and  $R_2$  in parallel (denoted as  $R_1 \parallel R_2$ ) is equal to their product divided by their sum:

$$R = \frac{R_1 R_2}{R_1 + R_2} \quad (1.7)$$

For  $n$  equal resistors in parallel, the equivalent resistance is equal to the value of an individual resistor  $R_1$  divided by  $n$ :

$$R_{TP} = \frac{R_1}{n} \quad (1.8)$$

The maximum voltage that can be applied across resistors in parallel is limited by the smallest value of RCWV for a resistor in the circuit.



**example 1.7** Three resistors of 120, 120, and 12  $\Omega$  are connected in parallel. The voltage across the parallel circuit is 240 V, and the RCWV of the 120- $\Omega$  resistors is 1000 V and 750 V for the 12- $\Omega$  resistor. Determine (a) the total equivalent resistance of the circuit, (b) the power dissipated in each resistor, (c) the maximum voltage that can be impressed across the parallel circuit. If the three resistors are replaced by a single equivalent resistor, determine (d) its power rating.

**solution** (a) Considering the two 120- $\Omega$  resistors, by Eq. (1.8),  $R_{TP} = 120/2 = 60 \Omega$ . By Eq. (1.7), the equivalent parallel resistance of the circuit is  $60 \parallel 12 = (60 \times 12)/(60 + 12) = 10 \Omega$ .

(b) By Eq. (1.3c),  $P = E^2/R$ , we have

$$P_{120} = \frac{(240)^2}{120} = 480 \text{ W}$$

$$P_{12} = \frac{(240)^2}{12} = 4800 \text{ W}$$

(c) The maximum voltage, limited by the smaller value of RCWV, is 750 V.

(d) In a we found that the equivalent resistance is 10  $\Omega$ . The power rating, therefore, is  $(240)^2/10 = 5760 \text{ W}$ .

Examples 1.5 and 1.7 illustrate that a high power rating can be realized by connecting lower power-rated resistors in series or in parallel.

## 1.7 SPECIAL RESISTORS

A number of special resistors and resistive networks are available. Some of these components are considered in this section.

**HIGH-VOLTAGE RESISTORS** Resistors that operate at voltages as high as 40 000 V are available. In one type of construction, a carbon-film resistance element is vacuum-sealed in a glass envelope.

**HIGH-MEGOHM RESISTORS** Using a semiconductor glass for the resistance element, resistance values as high as 1 000 000 M $\Omega$  (a million, million ohms) are obtainable. Measuring in length in the order of 0.25 in. and about 0.05 in. in diameter, the application for high-megohm resistors includes radiation detectors, electrometers, and use in FET circuits.

**DIP NETWORKS** Resistors, in various configurations packaged in the *dual-in-line package* (DIP), used for integrated circuits, constitute a DIP network. The DIP contains typically 14 or 16 pins. An example of a DIP ladder network is illustrated in Fig. 1.22. Innumerable resistor configurations are available. The resistors are generally thin or thick film.

## 1.8 VARIABLE RESISTORS (POTS)

A variable resistor, commonly referred to as a potentiometer or pot, converts the rotation of a shaft to an output voltage. There are basically three types of pots:

1. Single turn
2. Multiple turn
3. Trimmer

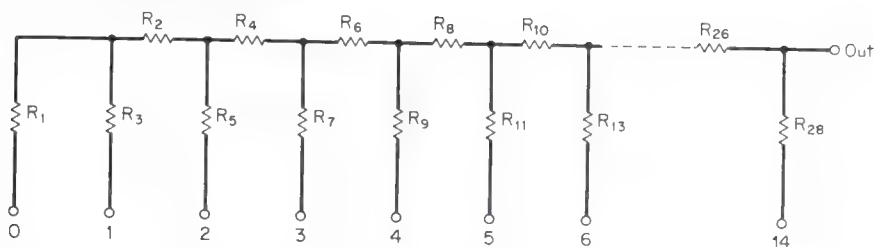
Examples of the three types of pots are illustrated in Figs. 1.23 to 1.25.

The single-turn pot may be regarded as the "workhorse" of variable resistors. Available in resistance ranges from 50  $\Omega$  to 5 M $\Omega$  and higher, in tolerances of  $\pm 10$  and  $\pm 20$  percent, and in power ratings of 2 and 3 W, the single-turn pot enjoys a wide spectrum of applications. It is used, for example, as a gain, treble, or base control in an amplifier and as the brightness and contrast controls in a TV receiver. In some applications, two or more pots sharing a common rotating shaft are required. Such a combination, referred to as a *ganged pot*, is illustrated in Fig. 1.26.

Multiple-turn pots are used in applications that require the precise setting of a resistance value. An example is the setting of coefficient pots in an analog computer.



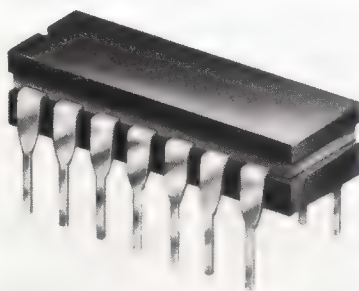
## 1-14 Characteristics of Resistors



Resistors:  $R_1, R_3, R_5, R_7, R_9, R_{11}, R_{13}, R_{15}, R_{17}, R_{19}, R_{21}, R_{23}, R_{25}, R_{27}, R_{28}$ ; each =  $2R$

Resistors:  $R_2, R_4, R_6, R_8, R_{10}, R_{12}, R_{14}, R_{16}, R_{18}, R_{20}, R_{22}, R_{24}, R_{26}$ ; each =  $R$

(a)

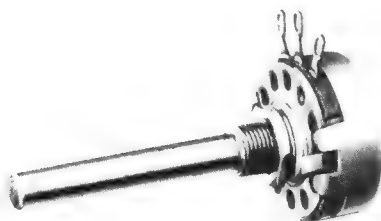


(b)

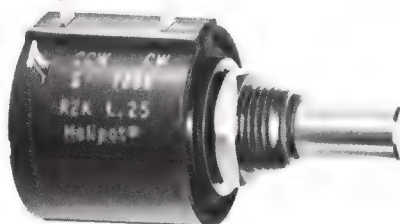
**Fig. 1.22** An example of a DIP network: (a) Ladder network. (b) DIP with 14 pins. (Courtesy RCL Electronics, Inc.)

These pots generally have 10 turns and are available in typical resistance ranges of  $50\ \Omega$  to  $250\ \text{k}\Omega$ , in tolerances of  $\pm 3$  percent, and in power ratings up to  $5\ \text{W}$ .

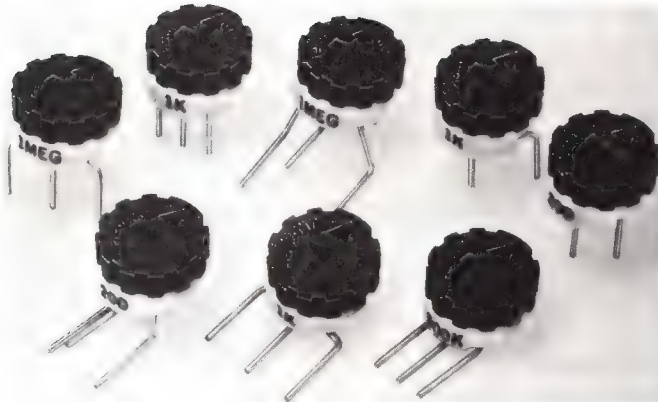
The trimmer pot, which may be viewed as a "set and forget me" pot, is used generally for a one-time adjustment of resistance. Its rotational life, compared to that of other pots, is therefore limited. Trimmer pots are available as single- and multiple-turn units. Typically, their resistance range is from a few ohms to  $5\ \text{M}\Omega$ , tolerance is  $\pm 10$  percent, and power rating is  $1\ \text{W}$ .



**Fig. 1.23** An example of a single-turn pot that enjoys a wide variety of application. (Courtesy Ohmite Manufacturing Company.)



**Fig. 1.24** An example of a multiple-turn pot. (Courtesy Beckman Instruments, Inc., Helipot Division)



**Fig. 1.25** Examples of trimmer pots. (Courtesy Beckman Instruments, Inc., Helipot Division)

**TYPES OF MATERIALS** Four basic types of materials are used in the construction of pots:

1. Carbon
2. Cermet
3. Conductive plastic
4. Wire-wound

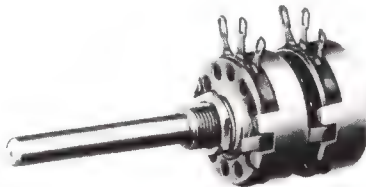
Pots other than wire-wound are referred to as non-wire-wound. A comparison of different types of pots is provided in Table 1.2.

**Carbon** The resistance element is made by spraying or brushing a carbon-resistance compound on an insulating material, such as laminated plastic. Its chief merit is its low cost. The carbon pot exhibits a high TCR and is susceptible to moisture, and its maximum operating temperature is generally limited to 85°C.

Variations of the carbon pot that exhibit somewhat improved characteristics are the *carbon-ceramic* and *molded-carbon* types. In the carbon-ceramic pot, the carbon compound is screened on a ceramic substrate. Molded-carbon pots are made by molding, simultaneously, the carbon-resistance element and other parts of the pot.

**Cermet** In the cermet pot, a mixture of precious metals and glass or ceramic powder is screened on a ceramic substrate. The TCR of a cermet pot may be as low as 50 ppm/°C and its operating temperature as high as 150°C. The cermet pot is very stable and can withstand overloads.

**Conductive plastic** One method of making a conductive plastic pot is by spraying, under pressure, a special liquid suspension on a plastic material, such as Mylar. The suspension contains carbon, a solvent, and a filler. The resistance value depends on the mix of carbon and filler. Conductive plastic pots exhibit good linearity and long rotational life. They are, however, susceptible to humidity.



**Fig. 1.26** A dual-ganged pot. (Courtesy Ohmite Manufacturing Company)

**Wire-wound** In the wire-wound pot, the resistance element is wound on a phenolic, plastic, or glass-filled laminate card. Low-resistance wire-wound pots are generally wound with a copper alloy wire and high-resistance pots with a nickel-chromium wire. The wire-wound pot exhibits the lowest TCR (20 ppm/°C), operates at temperatures as high as 150°C, and is available in high-power ratings. Because of its construction, the wire-wound pot has appreciable stray inductance and capacitance which may be a problem at high-frequency operation.

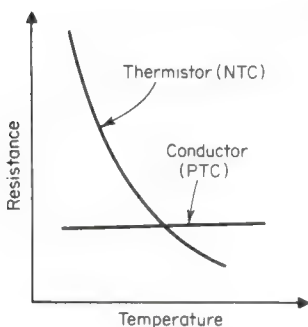
**RHEOSTAT** A wire-wound pot that can dissipate 5 and more watts is often referred to as a *rheostat*. The resistance wire is wound on an open ring of ceramic which is covered with vitreous enamel, except for the track of the wiper arm. Rheostats are used to control motor speed, x-ray tube voltages, welding current, ovens, and in many other high-power applications.

## 1.9 CHIP RESISTORS

A variety of thin- and thick-film resistor chips are available for hybrid microelectronic circuits (see Chap. 9). They measure as little as 30 by 30 mils (0.03 by 0.03 in.) and have resistance values from a few ohms to 1000 MΩ. Typical power rating is 0.25 W, and TCR is in the order of 20 to 200 ppm/°C.

## 1.10 THERMISTORS

A thermistor is a *nonlinear resistor*, made of semiconductor material, that is extremely sensitive to changes in temperature. For a small change in body temperature of a



**Fig. 1.27** The variation in resistance of a conductor (PTC) and a thermistor (NTC) with temperature.



**Fig. 1.28** Electrical symbol for a thermistor.

thermistor, there is an appreciable change in its resistance. Whereas most conductors have a positive temperature coefficient (PTC), the thermistor can exhibit a positive or negative temperature coefficient (NTC). Of particular interest is thermistors having a negative temperature coefficient.

A comparison of a conductor and an NTC thermistor is illustrated in Fig. 1.27. For the thermistor, the resistance decreases rapidly for elevated temperatures. The conductor having a positive temperature coefficient, however, exhibits a small increase in resistance with rising temperatures.

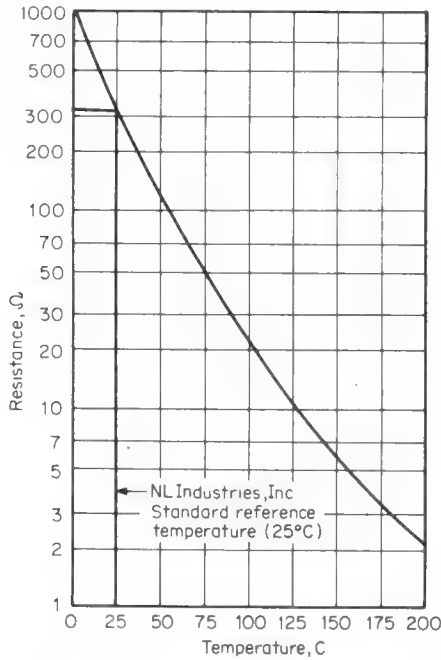
Thermistors are used in a wide variety of applications. These include the measurement and control of temperature, time delay, temperature compensation, and as liquid-level indicators. Thermistors are available as disk, washer, bead, and bolted assembly packages. The electrical symbol for a thermistor is illustrated in Fig. 1.28.

**CHARACTERISTIC CURVES AND PARAMETERS** In this section, typical characteristic curves and parameters for the thermistor are considered.

**Resistance-temperature characteristic** A typical resistance-temperature characteristic of a thermistor is illustrated in Fig. 1.29. Plotted is resistance as a function of

temperature. The resistance, for zero power dissipation, decreases rapidly with elevated temperature.

Three useful parameters for characterizing the thermistor are the time constant, dissipation constant, and resistance ratio. The time constant is the time for a thermistor to change its resistance by 63 percent of its initial value, for zero-power dissipation. Typical values of time constant range from 1 to 50 s.

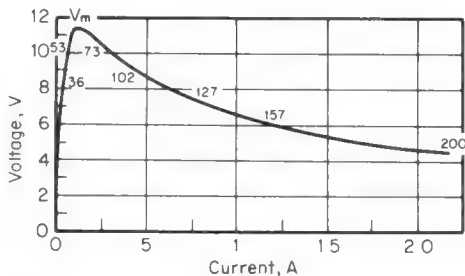


**Fig. 1.29** Typical resistance-temperature characteristic of a thermistor. (Courtesy N L Industries, Inc., Electronics Department)

The dissipation factor is the power necessary to increase the temperature of a thermistor by 1°C. Expressed in milliwatts per degree Celsius, typical values of dissipation factor are 1 to 10 mW/°C.

Resistance ratio is the ratio of the resistance at 25°C to that at 125°C. Its range is approximately 3 to 60.

**Static voltampere characteristics** A typical voltampere characteristic is shown in Fig. 1.30. Plotted is the voltage across the thermistor as a function of current flow. For



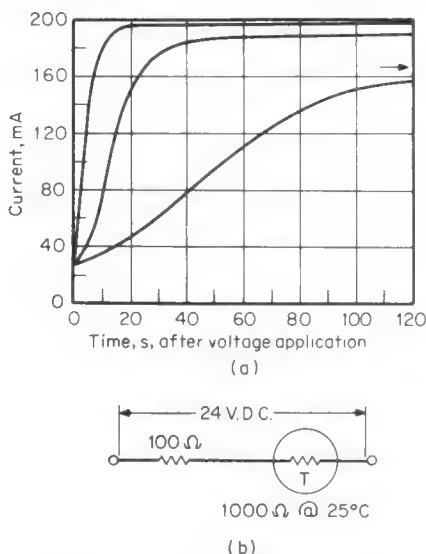
**Fig. 1.30** Typical static voltampere characteristic of a thermistor. (Courtesy N L Industries, Inc., Electronics Department)

## 1-18 Characteristics of Resistors

minute currents, the dissipated power is too small to heat the thermistor. In this region Ohm's law is obeyed. As the current is increased, the thermistor temperature rises and the resistance begins to decrease. At the *self-heating voltage*  $V_m$ , the characteristic begins to exhibit a negative slope. The temperature of the thermistor (in degrees Celsius) at various values of voltage and current is denoted by the numbers plotted along the curve.

**Current-time characteristic** Because of its mass, it takes a finite time for a current in a thermistor to reach its final value. This is illustrated in the current-time characteristic curves of Fig. 1.31a for three different thermistors.

The test circuit of Fig. 1.31b contains a 100- $\Omega$  resistor in series with a thermistor

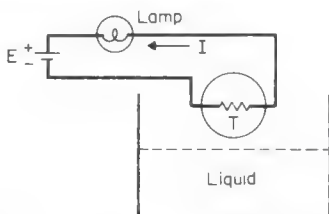


**Fig. 1.31** Current-time characteristics of a thermistor: (a) Typical family of curves. (b) Test circuit. (Courtesy N L Industries, Inc., Electronics Department)

whose resistance is 1000  $\Omega$  at 25°C. A voltage of 24 V dc is impressed across the circuit. For example, the thermistor having the characteristic of the top curve reaches its final value of approximately 200 mA in 20 s. The thermistor having the bottom characteristic, however, takes 120 s to approach its final value of 160 mA.

**example 1.8** Show how a thermistor may be used as a liquid-level indicator.

**solution** The circuit for a simple liquid-level indicator employing a thermistor is given in Fig. 1.32. Suspended above the liquid is a thermistor in series with an indicator, such as a lamp, and battery  $E$ . In air, the resistance of the thermistor is relatively low, thereby allowing sufficient current  $I$  to flow and light the lamp. When the liquid level reaches the thermistor, it

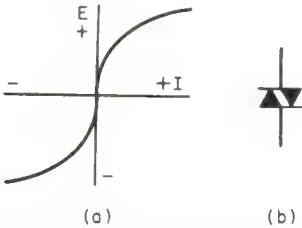


**Fig. 1.32** A simple liquid-level indicator using a thermistor.

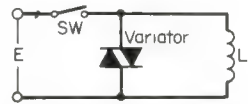
cools the device and its resistance is increased. Current  $I$  is reduced and the lamp is extinguished.

### 1.11 VARISTORS

The varistor, another example of a nonlinear resistor, is a device in which the current varies as a power of the impressed voltage. The resistance, therefore, is also reduced. In an ordinary resistor, the current is directly proportional to the impressed voltage and Ohm's law is obeyed. In the varistor, however, the current is proportional to the power of the impressed voltage  $E^n$ , where  $n$  is in the range of 2 to 6. A typical voltage-current characteristic curve and electrical symbol for the varistor are shown in Fig. 1.33.



**Fig. 1.33** The varistor: (a) Typical voltampere characteristic. (b) Electrical symbol. (Courtesy N L Industries, Inc., Electronics Department)



**Fig. 1.34** A simple surge-protection circuit using a varistor.

Applications for the varistor include voltage surge and protective circuits and the generation of nonsinusoidal waveforms. The varistor is made of silicon carbide and is available in disk, rod, and washer forms. It can withstand dc voltages as high as 10 000 V.

**example 1.9** Explain the operation of the surge protection circuit of Fig. 1.34.

**solution** Without the varistor connected across coil  $L$ , when the switch is opened, the high energy in the magnetic field develops a dangerously high voltage across  $L$ . This can result in the puncturing of the coil winding insulation. With a varistor in the circuit, the high voltage across it causes appreciable current to flow. The resistance of the device is thereby reduced and the magnetic field energy is dissipated in the varistor, instead of in the coil.





## Chapter 2

# Characteristics of Capacitors

### 2.1 INTRODUCTION

Next to resistors, capacitors are the most widely used passive element in circuits. They are available as fixed and variable units having capacitance values from a few picofarads (pF) to thousands of microfarads ( $\mu\text{F}$ ). For achieving unique characteristics, a large variety of materials are used in their construction. The abundance of different fixed capacitors available to the user is vividly portrayed in the photograph of Fig. 2.1. Because of such a plethora of units, the technician and engineer are often in a dilemma in selecting a capacitor for a given application.

**APPLICATIONS** The applications for capacitors may be broadly categorized as follows:

1. Blocking direct current. A capacitor cannot conduct direct current.
2. Coupling a signal from one circuit, or system, to another.
3. Bypassing a resistor to permit the easy flow of alternating current.
4. Filtering.
5. Tuning.
6. Generation of nonsinusoidal waveforms, such as a sawtooth wave.
7. Energy storage. Capacitors are used to build up sufficient electric charge to fire, for example, a flash tube or a laser.

### 2.2 GENERAL DESCRIPTION

The basic fixed capacitor consists of two metal plates, called *electrodes*, separated by an insulator, called a *dielectric*, as shown in Fig. 2.2a. In a variable capacitor, the relative position of one or more plates is changed with respect to a set of fixed plates, thereby varying the capacitance. Air is generally used as the dielectric for variable capacitors. Symbols for fixed and variable capacitors are shown in Fig. 2.2b and c, respectively.

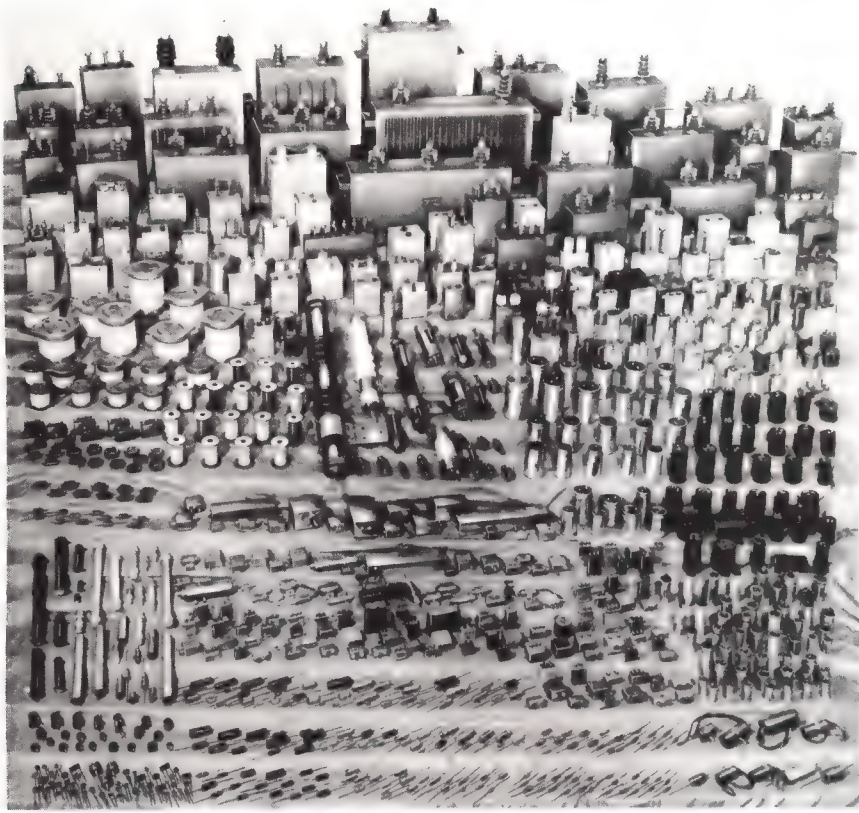
The capacitance of parallel-plate capacitors is directly proportional to the relative dielectric constant of the insulator and to the area of the plates. Further, the capacitance is greater if the separation between plates is small, and vice versa. These facts are conveniently expressed by the following equation:

$$C = \frac{k\epsilon_0 A}{d} \quad (2.1)$$

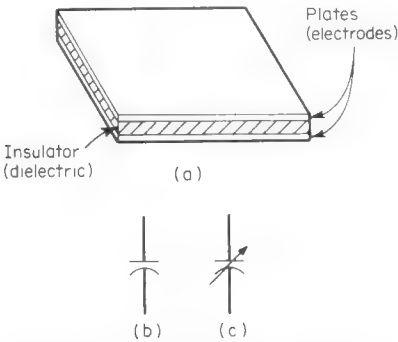
where  $C$  = capacitance in farads, F

$\epsilon_0$  = a constant, called the *permittivity of free space* and is equal to  $8.85 \times 10^{-12}$  F/m

$k$  = relative dielectric constant of insulator



**Fig. 2.1** A wide variety of types and styles of fixed capacitors are available to the technician and engineer. (Courtesy Sprague Electric Company)



**Fig. 2.2** The basic parallel-plate capacitor: (a) Construction. (b) Symbol for fixed capacitor. (c) Symbol for variable capacitor.

$A$  = area of plates,  $\text{m}^2$

$d$  = separation between plates,  $\text{m}$

From Eq. (2.1) it is seen that capacitance may be increased by increasing the area of the plates or the dielectric constant, and by decreasing the separation between plates.

**RELATIVE DIELECTRIC CONSTANT** The relative dielectric constant  $k$  compares different dielectrics with that of a vacuum for which  $k = 1$ . The dielectric constants for some commonly used materials are given in Table 2.1. Note that the dielectric constant of air is approximately equal to that of a vacuum.

**TABLE 2.1 Dielectric Constants of Some Commonly Used Materials**

Dielectric	$k$
Vacuum	1
Air	1.0006
Teflon	2
Polystyrene	2.5
Mylar	3
Paper, paraffin	4
Mica	5
Aluminum oxide	7
Tantalum oxide	25
Ceramic (low $k$ )	10
Ceramic (high $k$ )	100–10 000

**example 2.1** A parallel-plate capacitor has the following dimensions:  $A = 100 \text{ cm}^2$  and  $d = 1 \text{ mm}$ . Determine the values of capacitance if the dielectric used is: (a) air; (b) paper, paraffin; and (c) ceramic ( $k = 200$ ).

**solution** The given dimensions are first converted to meters. One square centimeter equals  $10^{-4} \text{ m}^2$ ; hence,  $100 \text{ cm}^2 = 100 \times 10^{-4} = 10^{-2} \text{ m}^2$ . One millimeter equals  $10^{-3} \text{ m}$ . Substitution of these values in Eq. (2.1) yields

$$C = k \times 8.5 \times 10^{-12} \times \frac{10^{-2}}{10^{-3}} = k \times 85 \times 10^{-12} \text{ F}$$

(a) The dielectric constant for air is approximately one. Hence,  $C_{\text{air}} = 85 \times 10^{-12} \times 1 = 85 \times 10^{-12} \text{ F} = 85 \text{ pF}$ , since  $10^{-12} \text{ F}$  is equal to  $1 \text{ pF}$ .

(b) For paper coated with paraffin, from Table 2.1,  $k = 4$ . Therefore,  $C_{\text{paper}} = 85 \times 10^{-12} \times 4 = 350 \times 10^{-12} \text{ F} = 350 \text{ pF}$ .

(c)  $C_{\text{ceramic}} = 85 \times 10^{-12} \times 200 = 17 \times 10^{-9} \text{ F} = 0.017 \mu\text{F}$ , since  $10^{-9} \text{ F}$  is equal to  $1 \mu\text{F}$ .

The particular dielectric used exerts a strong influence on the value of a capacitor. As we shall see, capacitors are designated by the dielectric used in their construction.

**CHARGE AND ENERGY** The charge  $Q$  (coulombs) stored in a capacitor equals the product of the capacitance  $C$  (farads) and the voltage  $E$  (volts) across the capacitor:

$$Q = CE \quad (2.2a)$$

Solving Eq. (2.2a) for  $E$  and  $C$ ,

$$E = \frac{Q}{C} \quad (2.2b)$$

$$C = \frac{Q}{E} \quad (2.2c)$$

From the preceding equations it is seen that to store a given charge, more voltage is needed for a small capacitor and less voltage for a large capacitor.

The maximum voltage that can be impressed across a capacitor, without causing irreparable damage, depends on the dielectric used and the separation  $d$  of the plates. For a given dielectric, the allowable maximum voltage increases as the separation between plates increases. From Eq. (2.1), as  $d$  grows large for a given plate area and

## 2-4 Characteristics of Capacitors

dielectric the capacitance is reduced. One may conclude that for a given dielectric and physical size, a capacitor with a high-voltage rating has less capacitance than one with a low-voltage rating.

A charged capacitor has *energy stored* in the electric field existing between its plates. The expression for stored energy is

$$w = \frac{Ce^2}{2} \quad (2.3)$$

where  $w$  = stored energy in Joules, J (1 J = 1 W-s)

$C$  = capacitance, F

$e$  = voltage across capacitor terminals, V

**example 2.2** Determine the energy stored in a 100- $\mu$ F capacitor if the impressed voltage is 1 kV.

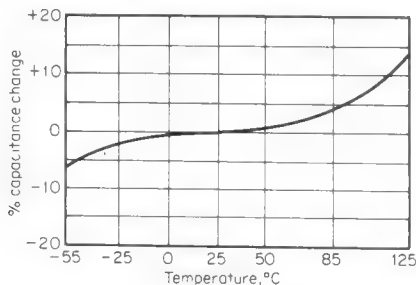
**solution** From Eq. (2.3)  $w = 100 \times 10^{-6} \times (1000)^2/2 = 50$  J.

## 2.3 CAPACITOR TERMS AND PARAMETERS

In this section commonly used terms and parameters for characterizing capacitors are defined. Where appropriate, typical curves showing the variation of parameters with temperature and frequency are included.

**CAPACITANCE** The basic unit for capacitance is the farad. Because practical capacitors have a capacitance much less than a farad, typical units employed are the microfarad and picofarad.

**AMBIENT TEMPERATURE** The actual value of capacitance depends on the temperature of the medium surrounding the capacitor. This is referred to as the *ambient temperature*. A typical curve showing the variation in capacitance as a function of temperature is given in Fig. 2.3. Note that *percent capacitance change* is plotted as a



**Fig. 2.3** A typical curve showing how capacitance varies with temperature. (Courtesy Electrocube)

function of temperature. At 25°C, which corresponds to room temperature, the percentage change is zero. This indicates that the changes in capacitance values are taken with respect to room temperature.

**example 2.3** A capacitor is rated at 100  $\mu$ F at 25°C. Using the curve of Fig. 2.3, determine the value of capacitance at -55 and +125°C.

**solution** From Fig. 2.3, at -55°C the change is approximately -6 percent. Hence, the value of  $C$  is  $100 - 100 \times 0.06 = 94$   $\mu$ F. At +125°C, the change is approximately +13 percent; the value of  $C$  at 125°C then is  $100 + 100 \times 0.13 = 113$   $\mu$ F.

**TOLERANCE** Tolerance is the variation in capacitance expressed as a percentage of its specified value at 25°C. The specified value at 25°C is referred to as the *nominal value*. For example, 20  $\mu$ F  $\pm 10\%$  means that the actual value is between  $20 - 20 \times 0.1 = 18$   $\mu$ F and  $20 + 20 \times 0.1 = 22$   $\mu$ F.

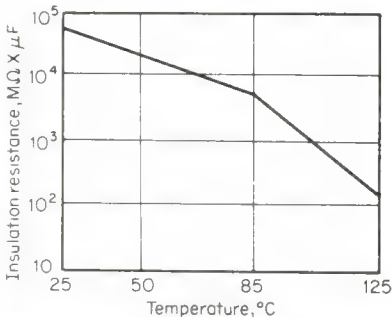
**TEMPERATURE COEFFICIENT** The temperature coefficient TC is the change in capacitance per degree change in temperature. It is generally expressed in parts per million per degree Celsius (ppm/°C). The temperature coefficient may be positive, negative, or zero. If the TC is positive, the letter *P* precedes the coefficient; if negative, *N* precedes the coefficient. The designation NPO is used for a zero-temperature coefficient capacitor.

**WORKING VOLTAGE** The working voltage is the maximum voltage that can be impressed across a capacitor for continuous operation. This rating must indicate whether the voltage is dc or ac; in general, the rating is not the same.

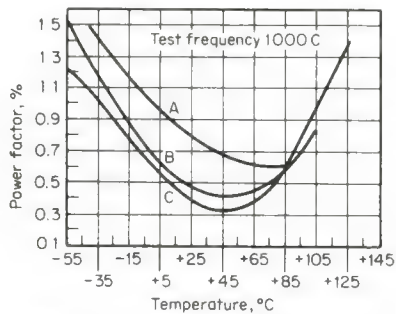
**BREAKDOWN VOLTAGE** The breakdown voltage is the maximum capacitor voltage that causes the dielectric to become damaged. It is also referred to as the *surge* and *test* voltage.

**DC LEAKAGE** Dc leakage refers to a minute direct current that flows in a capacitor at a specified direct voltage. Leakage is due to the presence of a few free carriers of charge in the dielectric. For this reason, a charge in a capacitor cannot be stored indefinitely, and it ultimately leaks off.

**INSULATION RESISTANCE** The insulation resistance *IR* is the resistance of the dielectric. The greater is the resistance, the less is the leakage current. Insulation resistance decreases with increasing temperature, as illustrated in Fig. 2.4.



**Fig. 2.4** The variation of insulation resistance with temperature. (Courtesy *Electrocube*)



**Fig. 2.5** Power factor as a function of temperature for three types of fixed capacitors. (a) Paper-dipped type (Type PD). (b) Mylar-paper-dipped (Type MPD). (c) Mylar. (Courtesy *The Electro Motive Manufacturing Company, Inc.*)

**CAPACITIVE REACTANCE** The capacitive reactance  $X_c$  (ohms) of a capacitor is equal to 0.159 divided by the product of the frequency (hertz) and capacitance (farads):

$$X_c = \frac{0.159}{fC} \quad (2.4)$$

The capacitive reactance increases as the frequency or capacitance decreases, and vice versa.

**POWER FACTOR** The power factor PF expresses the ratio of energy wasted to the energy stored in a capacitor. If, for example, PF = 2 percent, then 98 percent of the applied energy can serve a useful purpose and the remaining 2 percent is lost as heat in the capacitor. The variation of PF as a function of temperature is illustrated in Fig. 2.5. The power factor also increases with rising frequency.



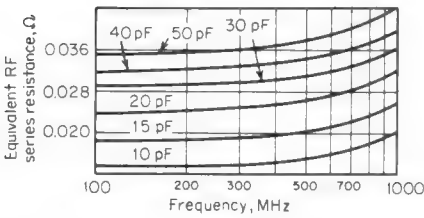
## 2-6 Characteristics of Capacitors

**EQUIVALENT SERIES RESISTANCE** Owing to the resistance of the capacitor plates and leads, the equivalent series resistance, ESR, is the net series resistance of a capacitor. An example of variation in ESR for capacitors designed to operate at UHF and microwave frequencies is shown in Fig. 2.6. The ESR increases with increasing frequency.

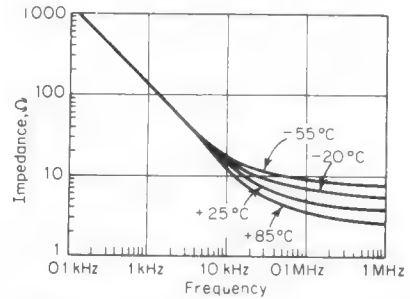
**IMPEDANCE** The impedance  $Z$  (ohms) of a capacitor is the total opposition to the flow of alternating current. It is equal to the square root of the sum of the capacitive reactance squared and the equivalent series resistance squared:

$$Z = \sqrt{X_C^2 + (\text{ESR})^2} \quad (2.5)$$

Typical curves of impedance as a function of frequency and temperature are illustrated in Fig. 2.7. Impedance of a capacitor decreases with increasing frequency and temperature. For a high-quality capacitor, the impedance is approximately equal to the capacitive reactance.



**Fig. 2-6** Variation in ESR with frequency for a porcelain capacitor designed to operate at ultrahigh and microwave frequencies. (Courtesy American Technical Ceramics)

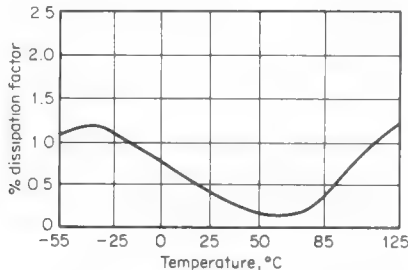


**Fig. 2.7** The impedance of a fixed capacitor decreases with increasing frequency and temperature. (Courtesy Sprague Electric Company)

**DISSIPATION FACTOR** The dissipation factor DF is another parameter used to describe the quality of a capacitor. The DF is expressed as a percentage and defined as the ratio of ESR to  $X_C$ :

$$\text{DF} = \frac{\text{ESR}}{X_C} \times 100\% \quad (2.6)$$

The lower the value of DF is, the better is the capacitor. A typical curve showing how DF varies with temperature is given in Fig. 2.8.



**Fig. 2.8** The variation of dissipation factor DF with temperature. (Courtesy Electro-cube)

**QUALITY FACTOR (Q)** The quality factor is one over the dissipation factor:

$$Q = \frac{1}{DF} = \frac{X_C}{ESR} \quad (2.7)$$

The greater the value of  $Q$  is, the better is the capacitor. Parameters  $DF$  and  $Q$  are relatively simple to determine on an impedance bridge.

**RIPPLE CURRENT AND VOLTAGE** Ripple current (voltage) is the ac component of a unidirectional current (voltage). Ripple is present, for example, in the output of a dc power supply. If the ripple component in a capacitor is less than 5 percent, it may be considered negligible. For values greater than 5 percent, the ripple contributes to the heating of the capacitor.

## 2.4 FIXED CAPACITORS

Although the types, shapes, and sizes of capacitors are nearly countless (see Fig. 2.1), the basic structure of Fig. 2.2, namely two plates separated by a dielectric, is common to all fixed capacitors. The plates and dielectric may be rolled in a tube, placed in a

**TABLE 2.2 Typical Characteristics of Commonly Used Fixed Capacitors**

Type	Capacitance range	Maximum working voltage, V	Maximum operating temperature, °C	Tolerance, %	Insulation resistance, MΩ
Mica	1 pF–0.1 μF	50 000	150	±0.25 to ±5	>100 000
Silvered mica	1 pF–0.1 μF	75 000	125	±1 to +20	1000
Paper	500 pF–50 μF	100 000	125	±10 to +20	100
Polystyrene	500 pF–10 μF	1000	85	±0.5	10 000
Polycarbonate	0.001–1 μF	600	140	±1	10 000
Polyester	5000 pF–10 μF	600	125	±10	10 000
Ceramic:					
Low $k$	1 pF–0.001 μF	6000	125	±5 to ±20	1000
High $k$	100 pF–2.2 μF	100	85	+100 to –20	100
Glass	10 pF–0.15 μF	6000	125	±1 to ±20	>100 000
Vacuum	1–5000 pF	60 000	85	±5	>100 000
Energy storage	0.5–250 μF	50 000	100	±10 to ±20	100
Electrolytic:					
Aluminum	1 μF–1 F	700	85	+100 to –20	<1
Tantalum	0.001–1000 μF	100	125	±5 to ±20	>1

disk or rectangular package, or otherwise to reduce the overall size. There are nearly 200 capacitor manufacturers in the United States. Their general thrust has been to develop larger capacitance values in smaller packages. The ratio of the capacitance value to the volume of the package is the *volumetric efficiency* of the capacitor.

Capacitors may be grouped into four broad categories:

1. Fixed
2. Variable
3. Chip
4. Voltage-variable (varactor)

Because the dielectric is a prime influence on capacitance size, a host of dielectrics are used in the construction of capacitors. Consequently, capacitors are classified according to their dielectrics. The characteristics of the most commonly used capacitors are summarized in Table 2.2.

**MICA** One of the earliest made, and still used, capacitors employs mica for the dielectric. Mica, a natural mineral, is a chemically inert and highly stable dielectric. The mica capacitor generally has a “sandwich” structure (Fig. 2.9), consisting of interleaving layers of tin-lead foil and mica. Mica capacitors are used over a wide tem-



## 2-8 Characteristics of Capacitors

perature range ( $-55$  to  $+150^{\circ}\text{C}$ ), and they have a high insulation resistance. Their capacitance values range from approximately  $1\text{ pF}$  to  $0.1\text{ }\mu\text{F}$ .

**SILVERED MICA** In the silvered mica a thin layer of silver, which is screened and fired onto the surface of the mica, is the conducting electrode. Among the advantages of a silvered mica over a mica capacitor are greater mechanical stability and more uniform characteristics.

**PAPER** The dielectric in paper capacitors is kraft paper impregnated with a wax or resin. It is generally packaged as a "rolled sandwich," shown in Fig. 2.10. Paper capacitors come in a large variety of values ( $500\text{ pF}$  to  $50\text{ }\mu\text{F}$ ), can operate in ambient temperatures as high as  $125^{\circ}\text{C}$ , are low cost, and can withstand high voltages. They are, however, temperature-sensitive. A  $10^{\circ}\text{C}$  rise in ambient temperature may reduce the life of the capacitor by as much as 50 percent.

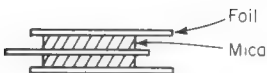


Fig. 2.9 The "sandwich" structure mica capacitor.

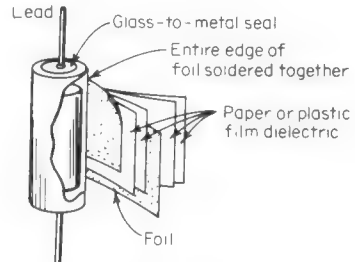


Fig. 2.10 The "rolled-sandwich" structure is commonly used for paper and plastic film capacitors.

The black band, usually printed on the package, is near the lead that is connected to the outer metal foil of the capacitor. To minimize noise, this lead should be connected to the lowest potential point in a circuit.

**PLASTIC FILM** Various plastics are used as the dielectrics of capacitors. Plastic capacitors are available in typical ranges of  $500\text{ pF}$  to  $10\text{ }\mu\text{F}$ . The plastics used include *polystyrene*, *polycarbonate*, and *polyester* (Mylar).

*Polystyrene capacitors* exhibit a low dissipation factor, small capacitance change with temperature, and very good stability. They tend to be large in size, and their maximum operating temperature is  $85^{\circ}\text{C}$ .

*Polycarbonate capacitors* have a dissipation factor and stability which approach those of polystyrene capacitors. They can operate at temperatures as high as  $140^{\circ}\text{C}$ .

*Polyester capacitors*, also referred to as Mylar capacitors, are the most widely used plastic capacitors. They are highly stable and exhibit a greater resistance to moisture than kraft paper. The maximum operating temperature of polyester capacitors is  $125^{\circ}\text{C}$ .

**CERAMIC** There are two basic types of ceramic dielectrics: low  $k$  and high  $k$ . Low- $k$  ceramic capacitors can be made to exhibit zero temperature coefficient, and they find wide use in temperature-compensation networks. The low- $k$  capacitor operates at voltages as high as  $6\text{ kV}$  and up to  $125^{\circ}\text{C}$ . Its capacitance value is generally limited to  $0.001\text{ }\mu\text{F}$ .

The high- $k$  ceramic capacitor has capacitance values up to  $2.2\text{ }\mu\text{F}$  and maximum working voltage of  $100\text{ V}$ . These capacitors, however, change their value appreciably with temperature, dc voltage, and frequency.

**GLASS** This capacitor is made by stacking alternate layers of aluminum foil and glass ribbon in a sandwich structure which is similar to the mica capacitor. These capacitors are extremely reliable and stable. It is claimed that any two glass capacitors, regardless of their value or size, exhibit a temperature coefficient within  $10\text{ ppm}/^{\circ}\text{C}$  of

each other. Glass capacitors can operate at voltages as high as 6000 V and up to 125°C. The range of capacitor values is 10 pF to 0.15  $\mu$ F.

**VACUUM** For very high voltages, a vacuum is often used as the dielectric. Vacuum capacitors range from 1 to 5000 pF and can operate at voltages as high as 60 kV.

**ENERGY STORAGE** Energy-storage capacitors store energy which can be discharged in a time interval from a fraction of a microsecond to several hundred microseconds. The dielectric used for energy storage capacitors often consists of a specially prepared and impregnated kraft paper. Capacitance values of 0.5 to 250  $\mu$ F and voltage ratings up to 50 kV are available.

**METALLIZED DIELECTRICS** If the dielectric, such as paper or plastic, is deposited with a thin metallic film instead of separate metal foils, volume savings up to 75 percent may be realized. Because of their low insulation resistance, however, these capacitors are not recommended for coupling and logic circuits.

**ELECTROLYTIC** In applications where a large value of capacitance in a small volume (high volumetric efficiency) is required, such as a filter in a power supply, the electrolytic capacitor is the choice. There are two basic types, *aluminum* and *tantalum* electrolytics, and are either *polarized* or *nonpolarized*.

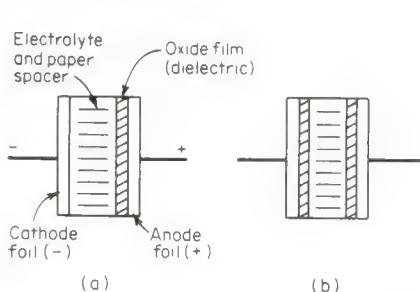


Fig. 2.11 An elementary electrolytic capacitor: (a) Polarized. (b) Nonpolarized.

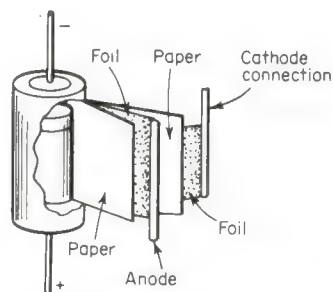


Fig. 2.12 The rolled-sandwich structure of an electrolytic foil capacitor.

In the polarized electrolytic capacitor, a plus sign is printed on the package near one of the two leads. When the capacitor is put in a circuit, the lead near the plus sign *must* be connected to a higher dc potential than the unmarked lead. If this is not done, the capacitor may be short-circuited and permanently damaged. No such restraint exists for connecting nonpolarized electrolytics in a circuit.

The basic structure of an electrolytic foil capacitor is illustrated in Fig. 2.11. For the polarized type of Fig. 2.11a, to increase its surface area, the aluminum (or tantalum) foil is etched. Then, a very thin layer of aluminum (or tantalum) oxide is electrochemically formed on the anode foil. The oxide layer becomes the dielectric for the capacitor. Because the oxide film is exceedingly thin and the surface area of the etched foil is large, the volumetric efficiency of electrolytic capacitors is high.

Separating the cathode and oxide-coated anode is a paper spacer which is soaked in an electrolyte solution. The spacer is required to prevent short circuiting between the cathode and anode foils. For high-voltage ratings, the spacer is thicker than for low-voltage units. The most common construction for electrolytic capacitors is the rolled sandwich, shown in Fig. 2.12.

The nonpolarized electrolytic capacitor in Fig. 2.11b has two oxide-coated anodes. For an equal-size polarized capacitor, the nonpolarized type has one-half the capacitance for the same voltage rating. Nonpolarized electrolytics should be used for such applications as ac motor starting, crossover networks, and large-pulse signals.

Aluminum electrolytic capacitors have high dc leakage and low insulation resistance.

## 2-10 Characteristics of Capacitors

Their shelf life is limited and their capacitance deteriorates with time and use. They enjoy, however, a high volumetric efficiency and are low in cost.

In addition to being very stable, tantalum oxide has nearly twice the dielectric constant of aluminum oxide. Tantalum electrolytic capacitors have a long shelf life, stable operating characteristics, increased operating temperature range, and a greater volumetric efficiency. The chief disadvantages of the tantalum, in comparison to the aluminum electrolytic, are its greater cost and lower voltage rating (see Table 2.2).

There are three types of tantalum electrolytics: *foil*, *wet anode*, and *solid anode*. The tantalum foil type is similar in construction to the aluminum foil electrolytic. The wet-anode capacitor is made by first molding, usually into the shape of a pellet, a mixture of tantalum powder and a binder. Under high temperature and in a vacuum, the pellet mixture is welded together (*sintered*), and the binder and impurities are driven off. The result is a porous pellet on which a layer of tantalum oxide is electrochemically formed. The volumetric efficiency of a wet-anode tantalum is about three times as great as the foil type.

The solid-anode tantalum is made by sintering an anode pellet on which is a layer of tantalum oxide. The pellet is then covered with a layer of manganese dioxide which serves as a solid electrolyte and cathode of the capacitor. This is followed by a layer of carbon and of silver paint, completing the cathode connection. The solid-anode construction is most widely used. It has the longest life and lowest leakage current of the three types of tantalum capacitors.

## 2.5 CONNECTING CAPACITORS

In this section we examine how fixed capacitors are connected in parallel and in series.

**CAPACITORS IN PARALLEL** One may think of connecting capacitors in parallel (Fig. 2.13) as effectively increasing the area of the plates. Because capacitance in-

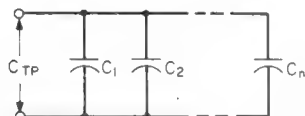


Fig. 2.13 Capacitors in parallel.

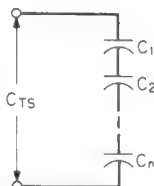


Fig. 2.14 Capacitors in series.

creases with plate area (more charge is stored in the capacitor), the total capacitance of capacitors connected in parallel,  $C_{TP}$ , is equal to the sum of the individual capacitances:

$$C_{TP} = C_1 + C_2 + \cdots + C_n \quad (2.8)$$

The working voltage of the parallel combination is limited by the smallest working voltage of the individual capacitors.

**example 2.4** In Fig. 2.13 assume that three capacitors are connected in parallel ( $n = 3$ ):  $C_1 = 40 \mu\text{F}$ , 300 V dc;  $C_2 = 25 \mu\text{F}$ , 450 V dc;  $C_3 = 80 \mu\text{F}$ , 200 V dc. Determine the capacitance and dc working voltage of the parallel combination.

**solution** The total capacitance  $C_{TP}$  is equal to the sum of  $C_1$ ,  $C_2$ , and  $C_3$ . Hence,  $C_{TP} = 40 + 25 + 80 = 145 \mu\text{F}$ . The dc working voltage is equal to the smallest rating of the individual capacitors. In this example, it is equal to the voltage rating of the 80- $\mu\text{F}$  capacitor, 200 V dc.

**CAPACITORS IN SERIES** Connecting capacitors in series may be thought of as increasing the separation of the outer plates of the combination (Fig. 2.14). The result is that the total capacitance  $C_{TS}$  is less than the smallest capacitance of the individual

capacitors. The relationship for  $C_{TS}$  is similar to that for resistors in parallel. For two capacitors in series,  $C_1$  and  $C_2$ , we have

$$C_{TS} = \frac{C_1 C_2}{C_1 + C_2} \quad (2.9)$$

If equal-valued capacitors are connected in series, the net value of capacitance equals the value of an individual capacitor divided by the number connected in series. For example, if three 150-pF capacitors are in series, the net capacitance is  $150/3 = 50$  pF.

The total dc working voltage rating of capacitors in series is equal to the sum of the dc ratings of the capacitors. This fact is sometimes used, for example, in the design of filters for high-voltage power supplies. For this purpose, electrolytics are generally chosen and connected in series to increase the dc working voltage. Because the insulation resistance of electrolytics is relatively low and variable with age and use, it is possible that one of the series capacitors will see a much greater voltage than it could withstand.

To prevent this from happening, *equalizing resistors*, of values between 20 and 50 k $\Omega$ , are placed across the capacitors (Fig. 2.15). The resistors form a voltage divider.

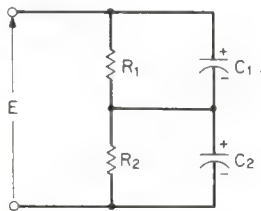


Fig. 2.15 Equalizing resistors  $R_1$  and  $R_2$  across electrolytic capacitors.

If they are equal, the voltage across each capacitor is  $E/2$  V. If they are unequal, the voltage across a capacitor is proportional to the resistance value.

## 2.6 COLOR CODING OF CAPACITORS

For large-size capacitors, the values of capacitance, working voltage, and tolerance are usually stamped on the packages. Color coding, based on the resistor code, is used for small capacitors. Figure 2.16 illustrates the coding schemes for various shapes of ceramic and mica capacitors.

## 2.7 VARIABLE CAPACITORS

The need for a variable capacitor arises in tuning circuits found in receivers, transmitters, and oscillators. A commonly used type is the *air-variable capacitor*. The capacitor of Fig. 2.17a is a *single-gang* type; the *dual-gang capacitor* of Fig. 2.17b is used in tuning, for example, a superheterodyne receiver.

Capacitance values for air-variable capacitors range from a few picofarads up to 500 pF; their maximum voltage rating is approximately 9 kV. For higher operating voltages (up to 60 kV), a *vacuum-variable capacitor* is used. A cross-sectional view of such a capacitor is illustrated in Fig. 2.18.

The air-variable capacitor is composed of two sets of plates, usually made of aluminum. One set, referred to as the *rotor*, is mounted on a shaft and meshes with a set of fixed metal plates, referred to as the *stator*. The stator and rotor plates are never permitted to touch each other.

Upon rotation of the shaft, either more or less area exists between the rotor and stator plates. Because capacitance is directly proportional to the area of the plates, the capacitance is thereby varied. The shape of the plates determines the manner in

## 2-12 Characteristics of Capacitors

<p>Current standard JAN and EIA code White (EIA) Black (JAN)</p>	<p>Ceramic capacitor codes (capacity given in pF)</p> <table> <tr> <th rowspan="2">Color</th><th rowspan="2">Digit</th><th rowspan="2">Multi-plier</th><th colspan="2">Tolerance</th><th rowspan="2">Temp-coeff ppm/°C</th><th colspan="2">Ext range temp-coeff</th></tr> <tr> <th>10pF or less</th><th>Over 10pF</th><th>Significant figure</th><th>Multiplier</th></tr> <tr><td>Black</td><td>0</td><td>1</td><td>2.0 pF</td><td>20%</td><td>33(N033)</td><td>0.0</td></tr> <tr><td>Brown</td><td>1</td><td>10</td><td>0.1 pF</td><td>1%</td><td>75(N075)</td><td>1.0</td></tr> <tr><td>Red</td><td>2</td><td>100</td><td></td><td>2%</td><td>150(N150)</td><td>1.5</td></tr> <tr><td>Orange</td><td>3</td><td>1000</td><td></td><td>2.5%</td><td>220(N220)</td><td>2.2</td></tr> <tr><td>Yellow</td><td>4</td><td>10000</td><td></td><td>5%</td><td>330(N330)</td><td>3.3</td></tr> <tr><td>Green</td><td>5</td><td></td><td>0.5 pF</td><td></td><td>470(N470)</td><td>4.7</td></tr> <tr><td>Blue</td><td>6</td><td></td><td></td><td></td><td>750(N750)</td><td>7.5</td></tr> <tr><td>Violet</td><td>7</td><td>0.01</td><td>0.25 pF</td><td></td><td>30(P030)</td><td></td></tr> <tr><td>Gray</td><td>8</td><td>0.1</td><td>10 pF</td><td>10%</td><td>General purpose bypass and coupling</td><td></td></tr> <tr><td>White</td><td>9</td><td>0.01</td><td></td><td></td><td>100(P100) (JAN)</td><td></td></tr> <tr><td>Silver</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>Gold</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table> <p>Ceramic capacitor voltage ratings are standard 500 V for some manufacturers 1000 V for other manufacturers unless otherwise specified</p>				Color	Digit	Multi-plier	Tolerance		Temp-coeff ppm/°C	Ext range temp-coeff		10pF or less	Over 10pF	Significant figure	Multiplier	Black	0	1	2.0 pF	20%	33(N033)	0.0	Brown	1	10	0.1 pF	1%	75(N075)	1.0	Red	2	100		2%	150(N150)	1.5	Orange	3	1000		2.5%	220(N220)	2.2	Yellow	4	10000		5%	330(N330)	3.3	Green	5		0.5 pF		470(N470)	4.7	Blue	6				750(N750)	7.5	Violet	7	0.01	0.25 pF		30(P030)		Gray	8	0.1	10 pF	10%	General purpose bypass and coupling		White	9	0.01			100(P100) (JAN)		Silver							Gold							<p>Molded flat paper capacitors (commercial code)</p>																								
Color	Digit	Multi-plier	Tolerance					Temp-coeff ppm/°C	Ext range temp-coeff																																																																																																																				
			10pF or less	Over 10pF	Significant figure	Multiplier																																																																																																																							
Black	0	1	2.0 pF	20%	33(N033)	0.0																																																																																																																							
Brown	1	10	0.1 pF	1%	75(N075)	1.0																																																																																																																							
Red	2	100		2%	150(N150)	1.5																																																																																																																							
Orange	3	1000		2.5%	220(N220)	2.2																																																																																																																							
Yellow	4	10000		5%	330(N330)	3.3																																																																																																																							
Green	5		0.5 pF		470(N470)	4.7																																																																																																																							
Blue	6				750(N750)	7.5																																																																																																																							
Violet	7	0.01	0.25 pF		30(P030)																																																																																																																								
Gray	8	0.1	10 pF	10%	General purpose bypass and coupling																																																																																																																								
White	9	0.01			100(P100) (JAN)																																																																																																																								
Silver																																																																																																																													
Gold																																																																																																																													
<p>Disk ceramics (5-dot system)</p>					<p>Molded flat paper capacitors (JAN code)</p>																																																																																																																								
<p>Disk ceramics</p>	<p>Typographically marked ceramics</p>	<table> <tr> <th rowspan="2">JAN letter</th><th colspan="2">Tolerance</th></tr> <tr> <th>10pF or less</th><th>Over 10pF</th></tr> <tr><td>C</td><td>0.25 pF</td><td></td></tr> <tr><td>D</td><td>0.5 pF</td><td></td></tr> <tr><td>F</td><td>1.0 pF</td><td>1%</td></tr> <tr><td>G</td><td>2.0 pF</td><td>2%</td></tr> <tr><td>J</td><td></td><td>5%</td></tr> <tr><td>K</td><td></td><td>10%</td></tr> <tr><td>M</td><td></td><td>20%</td></tr> </table>	JAN letter	Tolerance		10pF or less	Over 10pF	C	0.25 pF		D	0.5 pF		F	1.0 pF	1%	G	2.0 pF	2%	J		5%	K		10%	M		20%	<p>High-capacity tubular ceramics insulated or noninsulated</p>																																																																																																
JAN letter	Tolerance																																																																																																																												
	10pF or less	Over 10pF																																																																																																																											
C	0.25 pF																																																																																																																												
D	0.5 pF																																																																																																																												
F	1.0 pF	1%																																																																																																																											
G	2.0 pF	2%																																																																																																																											
J		5%																																																																																																																											
K		10%																																																																																																																											
M		20%																																																																																																																											
<p>Molded insulated axial lead ceramics</p>	<p>Button ceramics</p>	<p>Buttons silver mica</p>	<p>Temperature compensating tubular ceramics</p>																																																																																																																										
<p>Molded ceramics using standard resistor color code</p>	<p>Stand-off ceramics</p>	<p>Feed-thru ceramics</p>	<p>Extended-range T.C. tubular ceramics</p>																																																																																																																										
<p>Molded mica capacitor codes (Capacity given in pF)</p> <table> <tr> <th>Color</th><th>Digit</th><th>Multi-plier</th><th>Tolerance*</th><th>Class or characteristics</th></tr> <tr><td>Black</td><td>0</td><td>1</td><td>20%</td><td>A</td></tr> <tr><td>Brown</td><td>1</td><td>10</td><td>1%</td><td>B</td></tr> <tr><td>Red</td><td>2</td><td>100</td><td>2%</td><td>C</td></tr> <tr><td>Orange</td><td>3</td><td>1000</td><td>3%</td><td>D</td></tr> <tr><td>Yellow</td><td>4</td><td>10000</td><td></td><td>E (JAN)</td></tr> <tr><td>Green</td><td>5</td><td></td><td>5% (EIA)</td><td>F (JAN)</td></tr> <tr><td>Blue</td><td>6</td><td></td><td></td><td>G (JAN)</td></tr> <tr><td>Violet</td><td>7</td><td></td><td></td><td>I (EIA)</td></tr> <tr><td>Gray</td><td>8</td><td></td><td></td><td>J (EIA)</td></tr> <tr><td>White</td><td>9</td><td></td><td></td><td></td></tr> <tr><td>Gold</td><td></td><td>0.1</td><td>5% (JAN)</td><td></td></tr> <tr><td>Silver</td><td></td><td>0.01</td><td>10%</td><td></td></tr> </table> <p>Class or characteristic denotes specifications of design involving Q factors, temperature coefficients and production test requirement.</p> <p>All axial lead mica capacitors have a voltage rating of 300, 500 or 1000 V.</p> <p>* Or 10 pF whichever is greater.</p>		Color	Digit	Multi-plier	Tolerance*	Class or characteristics	Black	0	1	20%	A	Brown	1	10	1%	B	Red	2	100	2%	C	Orange	3	1000	3%	D	Yellow	4	10000		E (JAN)	Green	5		5% (EIA)	F (JAN)	Blue	6			G (JAN)	Violet	7			I (EIA)	Gray	8			J (EIA)	White	9				Gold		0.1	5% (JAN)		Silver		0.01	10%		<p>Molded paper capacitor codes (Capacity given in pF)</p> <table> <tr> <th>Color</th><th>Digit</th><th>Multi-plier</th><th>Tolerance</th></tr> <tr><td>Black</td><td>0</td><td>1</td><td>20%</td></tr> <tr><td>Brown</td><td>1</td><td>10</td><td></td></tr> <tr><td>Red</td><td>2</td><td>100</td><td></td></tr> <tr><td>Orange</td><td>3</td><td>1000</td><td></td></tr> <tr><td>Yellow</td><td>4</td><td>10000</td><td></td></tr> <tr><td>Green</td><td>5</td><td>100000</td><td>5%</td></tr> <tr><td>Blue</td><td>6</td><td>1000000</td><td></td></tr> <tr><td>Violet</td><td>7</td><td></td><td></td></tr> <tr><td>Gray</td><td>8</td><td></td><td></td></tr> <tr><td>White</td><td>9</td><td></td><td></td></tr> <tr><td>Gold</td><td></td><td></td><td>10%</td></tr> <tr><td>Silver</td><td></td><td></td><td>5%</td></tr> <tr><td>No color</td><td></td><td></td><td>20%</td></tr> </table>		Color	Digit	Multi-plier	Tolerance	Black	0	1	20%	Brown	1	10		Red	2	100		Orange	3	1000		Yellow	4	10000		Green	5	100000	5%	Blue	6	1000000		Violet	7			Gray	8			White	9			Gold			10%	Silver			5%	No color			20%	<p>Molded paper tubular</p>
Color	Digit	Multi-plier	Tolerance*	Class or characteristics																																																																																																																									
Black	0	1	20%	A																																																																																																																									
Brown	1	10	1%	B																																																																																																																									
Red	2	100	2%	C																																																																																																																									
Orange	3	1000	3%	D																																																																																																																									
Yellow	4	10000		E (JAN)																																																																																																																									
Green	5		5% (EIA)	F (JAN)																																																																																																																									
Blue	6			G (JAN)																																																																																																																									
Violet	7			I (EIA)																																																																																																																									
Gray	8			J (EIA)																																																																																																																									
White	9																																																																																																																												
Gold		0.1	5% (JAN)																																																																																																																										
Silver		0.01	10%																																																																																																																										
Color	Digit	Multi-plier	Tolerance																																																																																																																										
Black	0	1	20%																																																																																																																										
Brown	1	10																																																																																																																											
Red	2	100																																																																																																																											
Orange	3	1000																																																																																																																											
Yellow	4	10000																																																																																																																											
Green	5	100000	5%																																																																																																																										
Blue	6	1000000																																																																																																																											
Violet	7																																																																																																																												
Gray	8																																																																																																																												
White	9																																																																																																																												
Gold			10%																																																																																																																										
Silver			5%																																																																																																																										
No color			20%																																																																																																																										

Fig. 2.16 Color-coding schemes used for fixed capacitors.

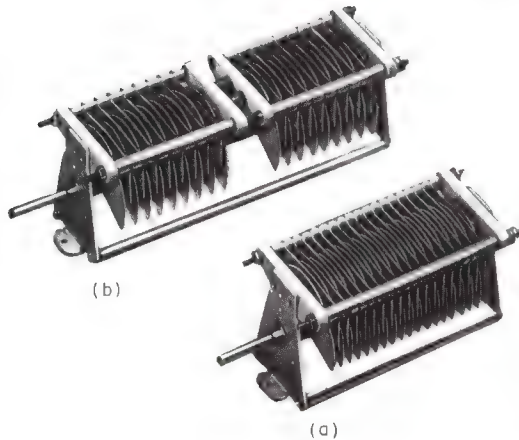
which capacitance varies with shaft rotation. For example, a linear or a nonlinear variation in capacitance with shaft position can be obtained.

**TRIMMERS** A variety of small variable capacitors using air, ceramic, mica, quartz, and other dielectrics are available for circuit applications. Referred to as *trimmer capacitors*, they are utilized for fine tuning and in hybrid microelectronic circuits. Their capacitance values range from a few picofarads to about 100 pF.

The capacitor of Fig. 2.19 is designed for hybrid circuits. It has a working voltage of 200 V dc, and in the 15- to 50-pF range, a Q of greater than 300 at 100 MHz. The ceramic trimmer of Fig. 2.20 has a working voltage of 250 V dc, and for the 4.5- to 50-pF range, a minimum Q of 100 at 100 MHz.

At high voltages and frequencies, the piston trimmer, with quartz or glass as the



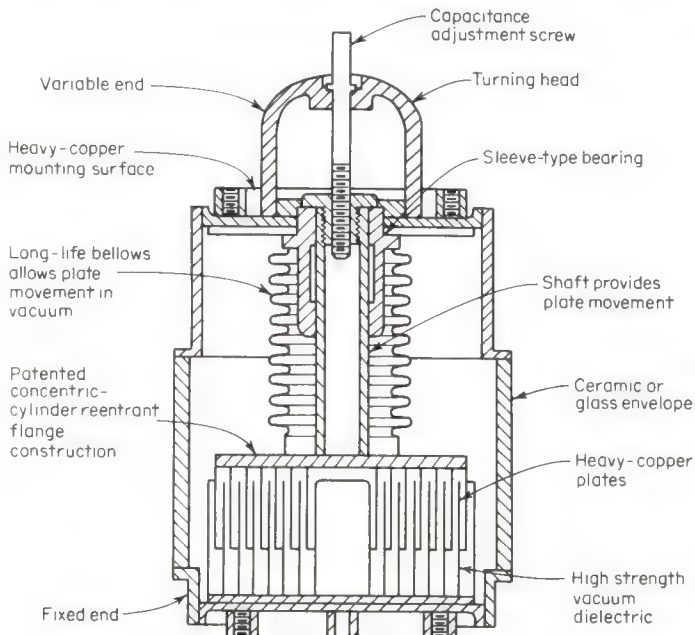


**Fig. 2.17** Examples of air-variable capacitors: (a) Single-gang. (b) Dual-gang. (Courtesy E. F. Johnson Company)

dielectric, is suitable (Fig. 2.21). A 1.5- to 38-pF piston trimmer with a quartz dielectric has a  $Q$  greater than 750 at 100 MHz and a working voltage of 1000 V dc.

## 2.8 CHIP CAPACITORS

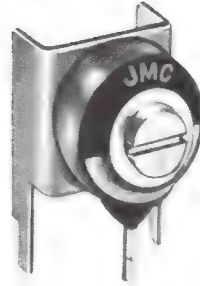
Often no larger than a match head, chip capacitors ranging in values from a few picofarads to 100  $\mu\text{F}$  are available for hybrid microcircuits. Because of their high volu-



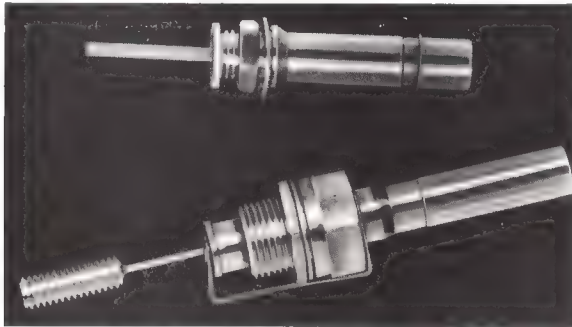
**Fig. 2.18** Construction details of a vacuum-variable capacitor. (Courtesy ITT Jennings)



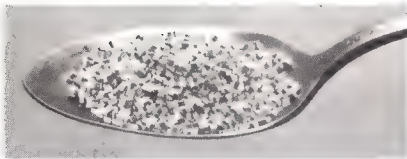
**Fig. 2.19** A high-Q ceramic tuning capacitor. It is used, for example, in electronic wrist-watches. (Courtesy Johanson Manufacturing Corp.)



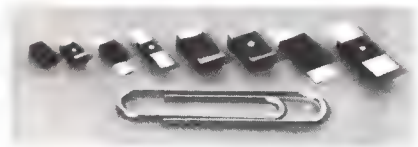
**Fig. 2.20** Side-tuned ceramic disk trimmer capacitor. (Courtesy Johanson Manufacturing Corp.)



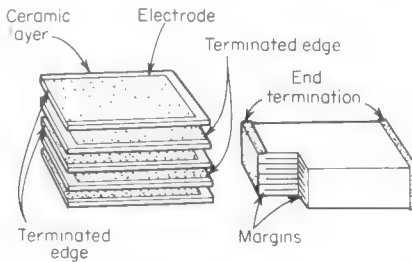
**Fig. 2.21** Examples of piston tuning capacitors. (Courtesy Johanson Manufacturing Corp.)



**Fig. 2.22** Ceramic chip capacitors. (Courtesy Sprague Electric Company)



**Fig. 2.23** Tantalum chip capacitors. (Courtesy Sprague Electric Company)



**Fig. 2.24** Construction features of a ceramic chip capacitor. (Courtesy West-Cap Division, San Fernando Electric Manufacturing Company)



metric efficiency, *ceramic* (Fig. 2.22) and *tantalum* (Fig. 2.23) chips are commonly used. Porcelain chip capacitors are employed in microwave circuits.

The ceramic chip capacitor is usually constructed as a multilayered sandwich of ceramic and screened-on conductor, illustrated in Fig. 2.24. In the porcelain chip capacitor, porcelain is used instead of ceramic for low loss at microwave frequencies. The tantalum chip capacitor is constructed in essentially the same manner as the solid-anode electrolytic.

A comparison of chip capacitors is provided in Table 2.3. They are all capable of operating at temperatures from  $-55$  to  $+125^{\circ}\text{C}$ . Using low- $k$  ceramic, chip capacitors up to approximately  $0.047\ \mu\text{F}$  with zero temperature coefficient are available. Although the maximum voltage rating of tantalum and porcelain chip capacitors is in the order of  $50\ \text{V}$ , this is no serious disadvantage. Transistors in monolithic and hybrid circuits generally operate at voltages that are well below  $50\ \text{V}$ . In terms of cost, tantalum chip capacitors tend to be the least expensive and porcelain chips the most expensive.

TABLE 2.3 Typical Characteristics of Chip Capacitors

Type	Capacitance range	Tolerance range, %	Maximum working voltage, V	Insulation resistance, M $\Omega$
Ceramic	10 pF–3.5 $\mu\text{F}$	$\pm 1$ to $\pm 20$	200	1000
Tantalum	100 pF–100 $\mu\text{F}$	$\pm 5$ to $\pm 20$	50	1500
Porcelain	1–330 pF	$\pm 1$ to $\pm 10$	50	100 000

## 2.9 VOLTAGE-VARIABLE CAPACITORS (VARACTORS)

Consider the pn junction diode of Fig. 2.25a. Separating the p and n regions is the depletion layer. When the pn junction is reverse-biased (the p region is negative with respect to the n region), the depletion region becomes devoid of free electrons and holes. One may therefore view the depletion layer of a reverse-biased diode as the dielectric, and the p and n regions as the plates of a capacitor.

As the reverse bias is increased, the depletion layer widens and the capacitance decreases, because the separation of the p and n regions has increased. If the reverse bias is reduced, the depletion layer contracts and the capacitance is increased. We thus have a diode whose capacitance varies inversely with the impressed reverse bias. The greater the magnitude of the reverse bias, the less the capacitance, and vice versa. Such a device is called a *voltage-variable capacitor*, or *varactor*. The symbol for this device is given in Fig. 2.25b.

Curves showing the variation in capacitance with reverse voltage for the varactor are illustrated in Fig. 2.26. Variation in capacitance is nonlinear. Over a range of

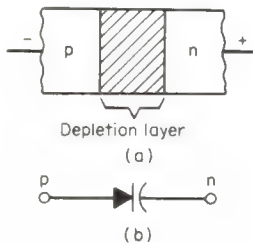


Fig. 2.25 Voltage-variable capacitor (varactor): (a) Reverse-biased p-n junction diode. (b) Symbol for voltage-variable capacitor.

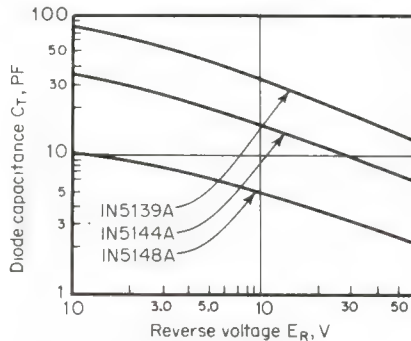
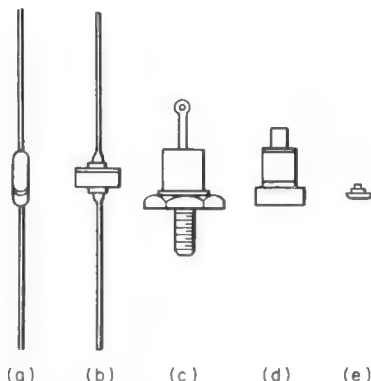


Fig. 2.26 Variations of varactor capacitance with reverse voltage. (Courtesy Codi Semiconductor)

## 2-16 Characteristics of Capacitors

reverse voltage from 1 to 60 V, for example, the variation in capacitance for the 1N5139A is approximately 80 to 14 pF. Diode capacitances as high as 2000 pF are possible for low-frequency applications. Microwave varactors may have a maximum capacitance of only 0.4 pF.

Applications for varactors are numerous and include TV tuners, high-power frequency multipliers, and frequency control. They can handle from 100 mW to hundreds of watts of r-f power. Their breakdown voltage ranges from approximately -6 to -300 V. Examples of various case styles used for varactors are illustrated in Fig. 2.27.



**Fig. 2.27** Examples of case styles for varactors: (a) General purpose. (b) High capacitance. (c) Power. (d) UHF power. (e) Microwave.

**PARAMETERS** Some of the significant parameters for characterizing the varactor will now be defined.

**Total diode capacitance ( $C_t$ )** The total diode capacitance  $C_t$  is the sum of the junction and case capacitances. It is generally specified at a reverse voltage  $E_R$  of -4 or -6 V. This parameter is important because it indicates whether the diode can be used at a particular frequency of interest. Variation in  $C_t$  with  $E_R$  is illustrated in Fig. 2.26.

**Series resistance ( $R_s$ )** Resistance  $R_s$  is the resistance in series with the junction of the diode. It varies with the reverse voltage, decreasing in value as the reverse voltage is increased.

**Figure of merit ( $Q$ )** The  $Q$  of a varactor is equal to

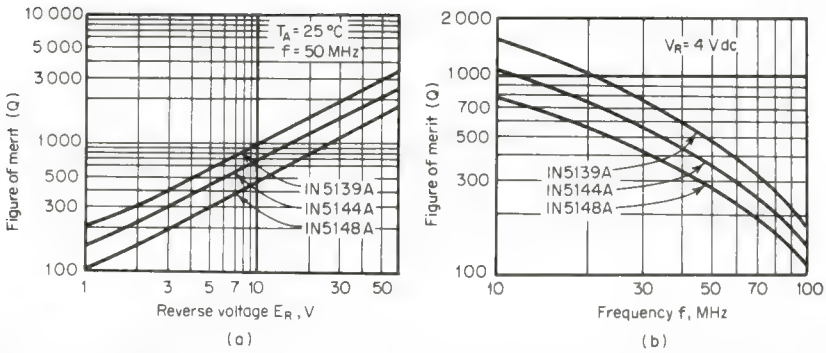
$$Q = \frac{0.159}{f R_s C_t} \quad (2.10)$$

where  $R_s$  is in ohms,  $f$  in hertz, and  $C_t$  in farads. The variation in  $Q$  with reverse voltage and frequency is illustrated in Fig. 2.28. As the reverse voltage increases, the values of  $R_s$  and  $C_t$  decrease, and  $Q$  rises. For rising frequencies, the losses in the diode increase and  $Q$  falls.

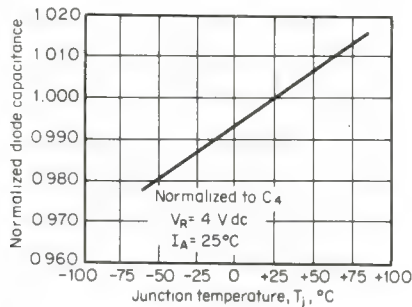
**Cutoff frequency ( $f_{co}$ )** Cutoff frequency  $f_{co}$  is the frequency where the  $Q$  of the device is unity.

**Variation of capacitance with temperature** The variation of  $C_t$  with junction temperature  $T_j$  is often presented as a normalized plot shown in Fig. 2.29. At a junction temperature corresponding to room temperature (25°C), the normalized value of  $C_t$  is one. As the temperature decreases, the capacitance falls and increases with increasing temperature.

**$C_t$  ratio** The ratio of capacitance variation at a reverse voltage of -4 or -6 V to the capacitance at approximately 80 percent of the breakdown voltage is referred to as the  $C_t$  ratio.



**Fig. 2.28** Variation in  $Q$  of a varactor with (a) reverse voltage and (b) frequency. (Courtesy Codi Semiconductor)



**Fig. 2.29** A normalized plot of varactor capacitance as a function of junction temperature. (Courtesy Codi Semiconductor)

**Conversion efficiency ( $\eta$ )** The conversion efficiency defines the performance of a varactor used as a frequency multiplier. This parameter is defined as the ratio of the output power  $P_o$  to the input power  $P_i$  and is expressed as a percentage:

$$\eta = \frac{P_o}{P_i} \times 100\% \quad (2.11)$$

**example 2.5** A tripler using a varactor multiplies an input frequency of 50 MHz to an output frequency of 150 MHz. If the r-f input power is 52.5 W and the output r-f power is 40 W, what is conversion efficiency of the tripler?

**solution** From Eq. (2.11)  $\eta = (40/52.5) \times 100\% = 76\%$ .



# Chapter 3

## Coils

### 3.1 INTRODUCTION

Capacitors and resistors are available in a wide variety of sizes and component values. It is possible to purchase almost any value, tolerance, or other rating in these components without special orders.

Coils, on the other hand, may not be as readily available. Technicians may find it necessary to wind their own coils when they want a specific value of inductance. This chapter gives the equations and characteristics of the most frequently used coils. A nomogram is included for simplifying the design of coils for specific applications.

### 3.2 TYPES OF FERROUS CORES AND THEIR EFFECTS IN COILS

When a core of magnetic material is inserted into an air-core coil, there is a great increase in coil inductance. A magnetic material is any kind of substance that is attracted by a magnet, and may be classified as being either hard or soft. The hard magnetic materials retain their magnetism after the magnetizing source has been removed, but the soft magnetic materials immediately lose magnetism upon removal of the magnetizing source. Saying that a magnetic material is soft does not mean that it is physically soft. A magnetic material may be physically extremely hard and brittle and yet be magnetically very soft. In other words, it very quickly loses all traces of magnetism as soon as the magnetizing source is removed. Soft magnetic materials are used as core material in radio-frequency and in audio-frequency coils and transformers.

**IRON-POWDER CORES** Iron-powder cores are composed of finely powdered iron or iron alloys mixed with a plastic binder. The cores are formed and pressed into final shapes at pressures ranging from 10 to 50 tons/in<sup>2</sup>, and then baked to set the binder.

Iron-powder cores are available in a large variety of different shapes and sizes and with different magnetic properties to meet different requirements.

**FERRITE CORES** Ferrites are also magnetic substances. They are unlike iron and steel and their various alloys in that ferrites are insulators. There are many different types of ferrites, and the exact chemical composition of each different type depends on its use; some ferrites that are useful at low frequencies cannot be employed at high frequencies and vice versa.

Table 3.1 gives the magnetic and physical properties of a number of different types of ferrites that are used as core material.

**LAMINATED CORES** The cores utilized in transformers and chokes that operate at power-line and audio frequencies are made of sheet steel laminations of various grades and thicknesses. These are treated in considerable detail in Chap. 5.

TABLE 3.1 Magnetic and Physical Properties of Ferrites Used as Core Material\*

Material	Initial permeability, $\mu_0$	Maximum permeability, $\mu_{\max}$	Flux density @ 14, B@H	Residual magnetism, Br	Coercive force, Hc	Frequency range, mHz	Loss factor, $\frac{1}{\mu_0 Q}$	Curie temperature, °C	Volume resistivity, $\Omega\text{-cm}$	Specific gravity, g/cm <sup>3</sup>
AM04	40	160	2200	400	4.0	10-80	@50 mHz $1.8 \times 10^{-4}$	500	$10^9$	4.3
AM12	125	340	2300	1000	1.9	0.2-20	@10 mHz $1.7 \times 10^{-4}$	300	$10^8$	4.4
AM20	200	640	2700	1800	1.5	0.05-8	@2.5 mHz $5 \times 10^{-5}$	280	$10^8$	4.4
AK04	400	1300	3500	1500	0.75	0.01-2	@100 kHz $2.5 \times 10^{-5}$	230	$10^3$	4.6
AK08	800	1500	3400	1500	0.45	0.01-1	@100 kHz $2.0 \times 10^{-5}$	200	$10^2$	4.6
AK16	1600	2800	3900	1000	0.20	1-400 kHz	@100 kHz $8.0 \times 10^{-6}$	140	$10^2$	4.5
AK20	2000	3400	3900	1800	0.16	1-250 kHz	@100 kHz $7.0 \times 10^{-6}$	135	$10^1$	4.6
AK30	3000	4500	4200	1500	0.18	1-250 kHz	@100 kHz $5.0 \times 10^{-6}$	200	$10^1$	4.6

\* Typical values as measured on toroidal cores.  
Courtesy The Arnold Engineering Company.



**Fig. 3.1** Ferrite tuning cores, toroids, special forms, and slugs.  
(Courtesy The Arnold Engineering Company)

**TUNING CORES, SLUGS, AND TOROIDS** When the core of a coil is adjustable within the coil, it is called a tuning core, or tuning slug. The nonadjustable cores are simply slugs or toroids. Ferrite tuning cores, toroids, special forms, and slugs of various sizes are shown in Fig. 3.1.

**CORE PERMEABILITY** Some idea of the increase in inductance achieved by the use of a core of magnetic material may be learned from a consideration of the permeability of the magnetic material—iron-powder, ferrite, or other. The permeability of air is always one, but the permeability of an iron-powder or ferrite core may be hundreds and even thousands of times greater than this. If the permeability of a magnetic core material is 100, its use increases the magnetic flux 100 times; this means that the inductance of the coil will also be increased 100 times.

Computations involving iron-powder and ferrite-core coils usually require technical data from core manufacturers. This will be discussed in a later paragraph.

### 3.3 TYPES OF NONFERROUS CORES

Besides cores of magnetic materials, cores of nonmagnetic metals (brass, copper, silver) are also sometimes employed at frequencies of about 50 mHz and above. Nonmagnetic cores have an effect on coil inductance just the opposite of that produced by a magnetic core; that is, they reduce, instead of increase, the inductance of a coil. When nonmagnetic cores are used, measures must be taken to reduce r-f core losses as much as possible. Since high-frequency currents flow mainly near the surface of a conductor, a thin metal sleeve or a silver-plated plastic core is sometimes utilized in the wire center. Solid nonmagnetic cores may be used when high coil efficiency is not a circuit requirement.

The effect of nonferrous cores on coil inductance is much less than that produced by cores of magnetic material. Nonferrous tuning cores are useful in equipment that requires a final adjustment of coil inductance before being put into service.

### 3.4 COMPUTING THE INDUCTANCE OF AIR-CORE COILS

A number of different formulas may be used for computing the inductance of air-core coils. Although these formulas take different forms, since they do give the same results, they may all be reduced to certain basic forms. The ones for the coil forms of Fig. 3.2a will give results sufficiently accurate for all practical purposes. The equations are listed here for convenience. Typical transmitter-type air-core coils are shown in Fig. 3.2b.

$$L = \frac{r^2 N^2}{9r + 10l} \quad (3.1)$$

$$L = \frac{0.8r^2 N^2}{6r + 9l + 10d} \quad (3.2)$$

$$L = \frac{r^2 N^2}{8r + 11d} \quad (3.3)$$



### 3-4 Coils

If coil dimensions are given in centimeters, instead of in inches, use the following:

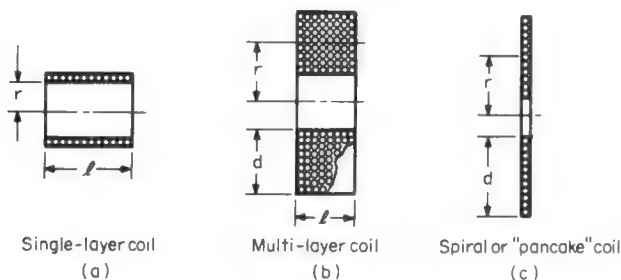
$$L = \frac{0.394r^2N^2}{9r + 10l} \quad (3.4)$$

$$L = \frac{0.315r^2N^2}{6r + 9l + 10d} \quad (3.5)$$

$$L = \frac{0.394r^2N^2}{8r + 11d} \quad (3.6)$$

The inductance of single-layer coils may also be determined by means of nomograms and by the use of special slide rules designed for this purpose.

The formulas of Fig. 3.2a all show that inductance is directly proportional to the square of the number of turns. Thus, if the number of turns in a coil is doubled, and



**Fig. 3.2** Computing the self-inductance of some typical air-core coils:

$$(a) L = \frac{r^2N^2}{9r + 10l} \quad (b) L = \frac{0.8r^2N^2}{6r + 9l + 10d} \quad (c) L = \frac{r^2N^2}{8r + 11d}$$

where  $L$  = coil inductance,  $\mu\text{H}$

$N$  = number of turns

$l$  = length of coil, inches

$d$  = depth of coil, inches

$r$  = mean radius of coil, inches

If coil dimensions are given in centimeters instead of inches, use the following:

$$(a) L = \frac{0.394r^2N^2}{9r + 10l} \quad (b) L = \frac{0.315r^2N^2}{6r + 9l + 10d} \quad (c) L = \frac{0.394r^2N^2}{8r + 11d}$$

coil dimensions are kept unchanged, inductance will be increased four times. This is strictly true only if there is very close coupling between turns, as there is between the turns and between the windings of iron-core transformers and chokes. However, for air-core coils, it is only approximately true.

**example 3.1** A single-layer coil like the one shown in Fig. 3.2a is to be wound in a space 2 in. long. How many turns must be wound in the 2-in. space if coil diameter is 1 in., and the required coil inductance is 250  $\mu\text{H}$ ?

**solution**

$$\begin{aligned} L &= \frac{r^2N^2}{9r + 10l} = 250 \\ &= \frac{(0.5)^2N^2}{9(0.5) + 10(2)} \end{aligned}$$

Solving for the number of turns,

$$\begin{aligned} N &= \frac{250[9(0.5) + 10(2)]}{(0.5)^2} \\ &= 156.5 \text{ turns} \end{aligned}$$

**example 3.2** A multilayer coil, such as that shown in Fig. 3.2*b*, is wound in a slot 1 in. square. Mean radius  $r$  of the coil is 2.25 in., and  $d = 1$ . If the coil contains 900 turns, what is the inductance of the coil?

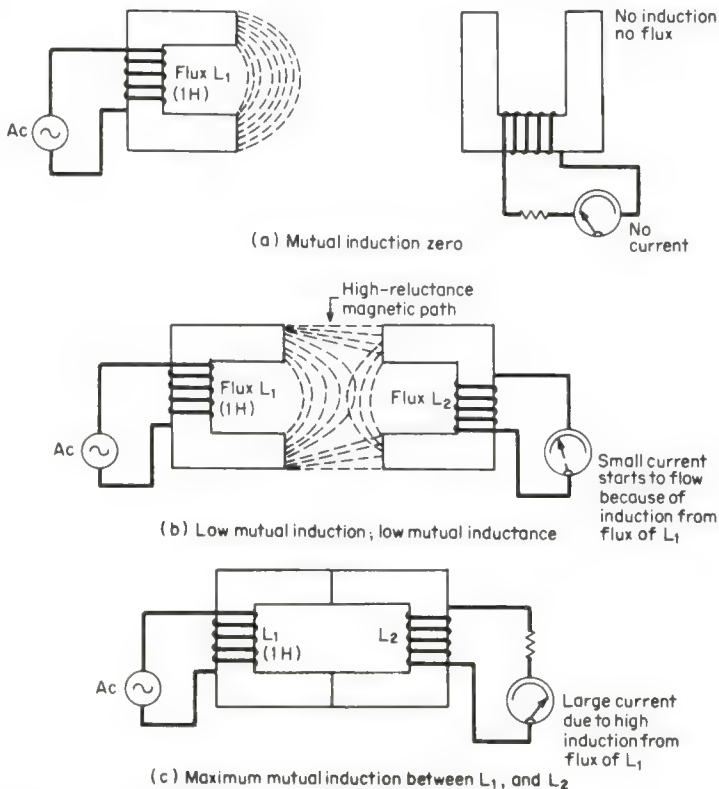
**solution**

$$\begin{aligned}
 L &= \frac{0.8r^2N^2}{6r + 9l + 10d} \\
 &= \frac{(0.8)(2.25)^2(900)^2}{6(2.25) + 9(1) + 10(1)} \\
 &= 100\,938 \mu\text{H} \\
 &= 100.938 \text{ mH}
 \end{aligned}$$

### 3.5 MUTUAL INDUCTANCE

Mutual induction, or the mutual inductive effect between two magnetically coupled coils, is due to the magnetic flux of the first coil acting on the turns of the second coil, as at the same time, the magnetic flux of the second coil acts on the turns of the first coil. To put it another way, there is a mutual interaction between the magnetic fields of the two coils.

*Mutual inductance* is a measure of the amount of mutual induction that exists between two magnetically coupled coils. Figure 3.3 illustrates the meaning of mutual induction. If the two coils are wound on separate cores and placed some distance from each other, as shown in Fig. 3.3*a*, there will be no magnetic coupling between the mag-



**Fig. 3.3** Mutual induction is the mutual interaction between the magnetic fluxes of  $L_1$  and  $L_2$ . The amount of mutual inductance depends on the coupling between  $L_1$  and  $L_2$ , which, in turn, depends on the reluctance of the magnetic circuit between  $L_1$  and  $L_2$  and on the self-inductances of the coils.

netic fields. Hence, the two coils will not have any mutually inductive effect on each other, and their mutual inductance will be zero.

If the two coils are brought nearer to each other, as in Fig. 3.3*b*, the magnetic field of coil  $L_1$  will start inducing a voltage in coil  $L_2$ , and a small current will flow in coil  $L_2$  which will then generate a magnetic field of its own because of the current. However, since the reluctance of the air path between the two coils will still be very high, the magnetic coupling between them will be very low, and they will have only a very low mutual inductance.

If the coils are next brought in close contact with each other, as in Fig. 3.3*c*, the magnetic coupling between the two coils will be a maximum; thus, mutual induction and mutual inductance between the two coils will also be a maximum.

It may be seen, therefore, that mutual inductance depends on the distance between two magnetically coupled coils. It also depends on the self-inductances of the coils and on the reluctance of the magnetic path between the coils. All these factors have an effect on the amount of coupling between the coils.

If we again consider the very closely coupled coils of Fig. 3.3*c*, and denote the inductance of coil  $L_1$  by  $L_1$  and that of coil  $L_2$  by  $L_2$ , then the combined inductance of the two coils connected in series aiding equals  $L_1 + L_2$  plus the mutual inductance due to the magnetic lines of force of  $L_1$  acting on the turns of  $L_2$ , plus the mutual inductance due to the magnetic lines of force of  $L_2$  acting on the turns of  $L_1$ . Since  $L_1$  and  $L_2$  are identical, their mutual induction effects are equal. Therefore, the total inductance is determined by the equation

$$L = L_1 + L_2 + 2M \quad (3.7)$$

where  $M$  = mutual inductance. Since  $L_1 = L_2$ ,

$$L = 2L_1 + 2M = 4L_1 = 4 \text{ H}$$

as determined by an actual measurement of the combined inductance of the two coils. Then, since  $L_1 = 1 \text{ H}$ ,

$$M = \frac{1}{2}(L - 2L_1)$$

$$M = \frac{1}{2}(4 - 2 \times 1) = 1 \text{ H}$$

In any other case in which two identical coils are as closely coupled as the two coils of Fig. 3.3*c*, mutual inductance will always be approximately the same as the inductance of either coil.

It should be noted that even if  $L_1$  and  $L_2$  are not equal to each other, it cannot be said that one of the coils has a mutual inductance that differs from the mutual inductance of the other. In fact, neither coil by itself has any mutual inductance. It is only when the two coils are magnetically coupled to each other that a mutual inductive effect exists between them, and it is never correct to say that one of the coils has a mutual inductance.

**COUPLED COILS IN SERIES AND IN PARALLEL** Not all coils will be as closely coupled magnetically as the two coils of Fig. 3.3*c*. For some applications this would not even be a desirable condition. In any case, it would be impossible to achieve with air-core coils or with any coils in which the coupling medium has a low permeability or high reluctance. Moreover, two coupled coils may have entirely different values of self-inductance. Hence in these cases the combined inductance of two coils connected in series aiding will be less than four times the inductance of one coil, even if the coupled coils are identical. The combined inductance of any two coils connected in series aiding and having self-inductances of  $L_1$  and  $L_2$ , respectively, will be given by

$$L = L_1 + L_2 + 2M \quad (3.8)$$

where  $L$  = combined inductance, and  $M$  = mutual inductance.

But if the coils are connected in series opposing, their combined inductance will be

$$L = L_1 + L_2 - 2M \quad (3.9)$$

And if two coils are connected in parallel with fields aiding,

$$L = \frac{1}{1/(L_1 + M) + 1/(L_2 + M)} \quad (3.10)$$

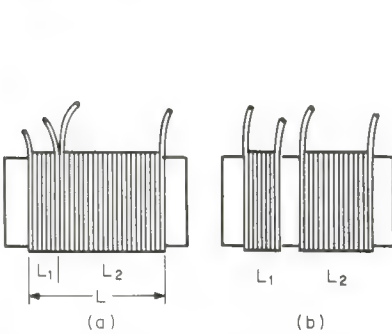
And for two coils in parallel with fields opposing,

$$L = \frac{1}{1/(L_1 - M) + 1/(L_2 - M)} \quad (3.11)$$

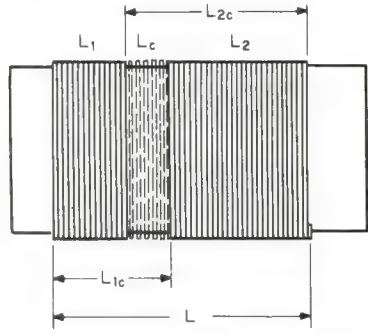
### COMPUTING THE MUTUAL INDUCTANCE OF SINGLE-LAYER AIR-CORE COILS

An important case for consideration is the mutual inductance between coils having air cores. Figure 3.4 shows two important examples, one with no spacing between the coils and the other with a space between them.

CASE 1: If there is no separation between the windings of two coupled coils wound on the same form, as shown in Fig. 3.4a, their combined inductance with the coils connected in series aiding is given by Eq. (3.8), and also by the equation for computing



**Fig. 3.4** Typical single-layer air-core coils: (a)  $L_1$  and  $L_2$  on the same form and touching; mutual inductance,  $M = \frac{1}{2}(L - L_1 - L_2)$ . (b) With  $L_1$  and  $L_2$  on the same form but separated by an air space, computing mutual inductance is more difficult.



**Fig. 3.5**  $L_1$  and  $L_2$  are single-layer coils wound on the same form.  $L_c$  is an assumed winding in the space between  $L_1$  and  $L_2$ . Mutual inductance between  $L_1$  and  $L_2$  is computed from the following formula:  $M = \frac{1}{2}(L + L_c - L_{1c} - L_{2c})$ .

the inductance of a single-layer coil [Eq. (3.1)]. Since the separate inductances of the two coils may also each be computed by using Eq. (3.1), the only unknown in Eq. (3.8) will be  $M$ . This can be determined from the equation

$$M = \frac{1}{2}(L - L_1 - L_2) \quad (3.12)$$

CASE 2: When the windings of two coupled coils are separated by an air space, as in Fig. 3.4b, imagine that the space between the coils is also wound with wire that has the same size as the windings of  $L_1$  and  $L_2$ . The fictitious center winding is called  $L_c$ . Also, assume that  $L_1$ ,  $L_c$ , and  $L_2$  are connected in series aiding, as shown in Fig. 3.5, and their combined inductance is found by using Eq. (3.1). Using the same equation, the inductance of  $L_1$  in series with the imaginary center winding  $L_c$  is found, then of  $L_2$  in series with  $L_c$ , and of  $L_c$  separately.  $L$ , the combined inductance of  $L_1$ ,  $L_c$ , and  $L_2$ , is then given by

$$L = L_{1c} + L_{2c} - L_c + 2M$$

Then

$$2M = L + L_c - L_{1c} - L_{2c}$$

and

$$M = \frac{1}{2}(L + L_c - L_{1c} - L_{2c}) \quad (3.13)$$

**example 3.3** Two single-layer coils,  $L_1$  and  $L_2$ , are wound side by side on the same form, as in Fig. 3.4a. Coil  $L_1$  has 40 turns and  $L_2$  has 60. The total winding space is 2 in long, and coil radius  $r$  is 0.5 in. Compute:

- Total inductance with the coils connected in series aiding
- Mutual inductance
- Total inductance with the coils connected series opposing
- Total inductance with  $L_1$  and  $L_2$  connected in parallel and fields *aiding*
- Total inductance of  $L_1$  and  $L_2$  in parallel with fields opposing

### 3-8 Coils

**solution** (a) Using Eq. (3.1),

$$\begin{aligned} L &= \frac{(0.5)^2 100^2}{9(0.5) + 10(2)} \\ &= \frac{0.25(10\,000)}{4.5 + 20} = 102.04 \mu\text{H} \end{aligned}$$

(b) Since  $L_1$  has 40 turns, and the total number of turns is 100 in a winding space 2 in long, the length of  $L_1$  is 0.4 of 2, or 0.8 in. Therefore,

$$L_1 = \frac{0.5^2(40)^2}{9(0.5) + 10(0.8)} = 32 \mu\text{H}$$

and

$$L_2 = \frac{0.5^2(60)^2}{9(0.5) + 10(1.2)} = 54.54 \mu\text{H}$$

Since the series-aiding inductance,

$$L = 102.08 = 32 + 54.54 + 2M$$

it follows that

$$M = \frac{1}{2}(102.08 - 32 - 54.54) = 7.77 \mu\text{H}$$

(c) With the coils connected in series opposing,

$$\begin{aligned} L &= 32 + 54.54 - 2M \\ &= 86.54 - 2(7.77) = 71 \mu\text{H} \end{aligned}$$

(d) With the coils connected in parallel and fields *aiding*,

$$\begin{aligned} L &= \frac{1}{1/(L_1 + M) + 1/(L_2 + M)} \\ &= \frac{1}{1/(32 + 7.77) + 1/(54.54 + 7.77)} \\ &= 24.28 \mu\text{H} \end{aligned} \quad (3.10)$$

(e) When the coils are connected in parallel with fields *opposing*,

$$\begin{aligned} L &= \frac{1}{1/(L_1 - M) + 1/(L_2 - M)} \\ &= \frac{1}{1/(32 - 7.77) + 1/(54.54 - 7.77)} \\ &= 15.96 \mu\text{H} \end{aligned} \quad (3.11)$$

**example 3.4** Two single-layer coils,  $L_1$  and  $L_2$ , are wound on the same form, as in Fig 3.4b. The coils are separated by an air space of  $\frac{1}{4}$  in, and coil radius,  $r$  is 1 in. Coil  $L_1$  contains 24 turns and is 1 in long;  $L_2$  contains 48 turns and is 2 in long. Compute:

(a) Mutual inductance

(b) Total inductance of  $L_1$  and  $L_2$  connected in series aiding

(c) Total inductance of  $L_1$  and  $L_2$  connected in parallel with fields *aiding*.

**solution** To determine the mutual inductance, the four inductances,  $L$ ,  $L_{1c}$ ,  $L_{2c}$ , and  $L_r$  must first be computed. Since  $L_1$  contains 24 turns and is 1 in long, the  $\frac{1}{4}$ -in air space would contain six turns; total number of turns in the  $3\frac{1}{4}$ -in winding space would then be  $24 + 6 + 48 = 78$  turns, and for 78 turns,

$$\begin{aligned} L &= \frac{1(78)^2}{9(1)10(3\frac{1}{4})} = 146.6 \mu\text{H} \\ L_{1c} &= \frac{1^2(30)^2}{9(1) + 10(1\frac{1}{4})} = 41.86 \mu\text{H} \\ L_{2c} &= \frac{1^2(54)^2}{9(1) + 10(2\frac{1}{4})} = 92.57 \mu\text{H} \\ L_r &= \frac{1^2(6)^2}{9(1) + 10(\frac{1}{4})} = 3.13 \mu\text{H} \end{aligned}$$

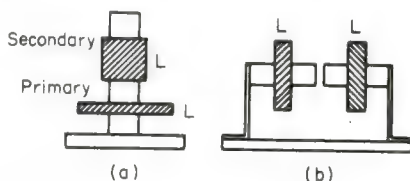
(a) The mutual inductance,

$$M = \frac{1}{2}(L + L_r - L_{1c} - L_{2c}) = \frac{1}{2}(146.6 + 3.13 - 41.86 - 92.57) = 7.65 \mu\text{H}$$

$$\begin{aligned}
 (b) \quad L_1 &= \frac{1^2(24)^2}{9(1) + 10(1)} = 30.32 \mu\text{H} \\
 L_2 &= \frac{1^2(48)^2}{9(1) + 10(2)} = 79.45 \mu\text{H} \\
 L &= L_1 + L_2 + 2M \\
 &= 30.32 + 79.45 + 2(7.65) = 125.07 \mu\text{H} \\
 (c) \quad L &= \frac{1}{1/(L_1 + M) + 1/(L_2 + M)} \\
 &= \frac{1}{1/(30.32 + 7.65) + 1/(79.45 + 7.65)} \\
 &= 26.44 \mu\text{H}
 \end{aligned}$$

### 3.6 COMPUTING THE VALUE OF INDUCTANCE FROM MEASUREMENTS

In some cases it is very difficult to determine mutual inductance by calculation. Two examples are shown in Fig. 3.6.



**Fig. 3.6** (a) Two multilayer coils on the same form. (b) Ferrite core r-f chokes separated by an air gap. Mutual inductance is computed from  $M = \frac{1}{4}(L_A - L_o)$ , where  $M$  = mutual inductance,  $L_A$  = total inductance of  $L_1$  and  $L_2$  in series aiding,  $L_o$  = total inductance of  $L_1$  and  $L_2$  in series opposing.

The two r-f coils of Fig. 3.6a are wound on the same iron-powder core but separated from each other. Figure 3.6b shows two r-f chokes wound on separate ferrite cores and separated from each other by a small air gap. In both of these, and in many similar instances, it would be extremely difficult to calculate mutual inductance from such known factors as the self-inductance of the coils. It is possible, however, to measure the combined inductance of the coils connected in series aiding and then in series opposing, denoting the series-aiding inductance by  $L_A$  and the series-opposing inductance by  $L_o$ . Two simultaneous equations are then formed.

$$\begin{aligned}
 L_A &= L_1 + L_2 + 2M \\
 L_o &= L_1 + L_2 - 2M
 \end{aligned}$$

Subtracting the second equation from the first,

$$L_A - L_o = 4M$$

from which

$$M = \frac{1}{4}(L_A - L_o) \quad (3.14)$$

where  $M$  = mutual inductance

$L_A$  = combined inductance of  $L_1$  and  $L_2$  connected in series aiding

$L_o$  = combined inductance of  $L_1$  and  $L_2$  connected in series opposing

**example 3.5** Two radio-frequency coils are separated by a small air gap, as shown in Fig. 3.6b. When the coils are connected in series aiding, their combined inductance, as measured on an inductance bridge, is 17 mH, and when the coils are connected in series opposing, combined inductance is 3 mH. What is the mutual inductance?

**solution**

$$M = \frac{17 - 3}{4} = 3.5 \text{ mH}$$



### 3.7 THE COEFFICIENT OF COUPLING

The coefficient of coupling, or coupling coefficient, is a measure of the amount of coupling between two coils; it shows whether the coupling is "tight" or "loose." The greatest possible value of the coefficient of coupling is one, in which case the *coupling* may be said to be 100 percent. This is very tight coupling. Coupling coefficient may be computed from the equation

$$k = \frac{M}{\sqrt{L_1 L_2}} \quad (3.15)$$

where  $k$  = coefficient of coupling (There are no units for  $k$ )

$M$  = mutual inductance in the same units as  $L_1$  and  $L_2$

$L_1, L_2$  = self-inductances of the separate coils

A coupling coefficient of 1, or 100 percent coupling, is approached only in well-designed iron-core transformers and chokes. In air-core transformers and coils it is very much less, but increases in transformers with iron-powder and ferrite cores.

**example 3.6** When the primary of an r-f antenna coil, such as that shown in Fig. 3.6a, is connected to the secondary in series aiding, the measured inductance is 3085  $\mu\text{H}$ ; when the primary and secondary are in series opposing, the measured inductance is 1405  $\mu\text{H}$ . If primary and secondary self-inductances are 2000 and 245  $\mu\text{H}$ , respectively, what is the coefficient of coupling between the primary and secondary?

**solution**

$$\begin{aligned} M &= \frac{1}{4}(L_A - L_0) \\ &= \frac{1}{4}(3085 - 1405) = 420 \mu\text{H} \end{aligned}$$

and the coefficient of coupling,

$$k = \frac{M}{\sqrt{L_1 L_2}} = \frac{420}{\sqrt{2000(245)}} = 0.6$$

### 3.8 COIL AND CONDUCTOR RESISTANCE AT RADIO FREQUENCIES

The resistance that a conductor offers to the flow of high-frequency alternating current through it is greater than the resistance of the same conductor to direct current. This is due to a characteristic of coils and conductors that is known as the skin effect. It is a characteristic of conductors that carry alternating current, but since it is a magnetic effect, it is much more pronounced in coils than in straight conductors.

Skin effect causes the current in a conductor to be much denser near the surface of the conductor than at its center, and is a result of the magnetic lines of force associated with any current-carrying conductor. These magnetic lines of force, or flux lines, not only surround the conductor but also exist in the interior of the conductor and are densest at its center. Consequently, as the instantaneous value of the alternating current changes, the inductance of a conductor in its interior is much greater than the inductance near its surface. The result is that more current flows near the surface, or "skin," of the conductor than in its interior. Since inductive effects increase with increasing frequency, skin effect also increases as frequency is increased.

**EDDY CURRENTS IN CONDUCTORS** Besides the skin effect, every conductor that carries alternating current also has eddy currents induced in it by its own changing magnetic field—just as a changing magnetic field induces eddy currents in the core material of a transformer or choke. These eddy currents constitute an additional loss which causes the effective ac or r-f resistance of a conductor (or coil) to be increased.

At very high frequencies, because of the skin effect and induced eddy currents in a conductor, a hollow, thin-walled tube is really a very much better conductor than a solid conductor of the same diameter. Some short-wave (high-frequency) coils are actually made with copper tubing, instead of with solid wire.

Since the resistance of a coil at ac frequencies depends on the frequency, in all ac problems this resistance must be considered at the operating frequency of the coil.

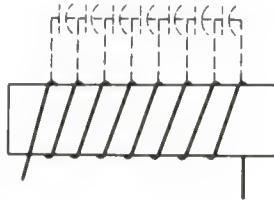
**DIELECTRIC LOSSES** Another source of loss in coils at high frequencies is that which occurs in the forms on which the coils are wound and in the insulating covering



of the wire. Since both of these are dielectric materials, any loss of this sort is called a dielectric loss. Even though a coil form may be made of the highest quality dielectric material, such as molded mica, steatite, resinite, or polystyrene, there is no perfect insulator, and dielectric losses cannot be entirely avoided at very high frequencies.

Dielectric loss, like skin effect and the eddy current loss in a conductor, also tends to increase the effective resistance of a coil at high frequencies.

**DISTRIBUTED CAPACITANCE** Since every turn of a coil is separated from its adjacent turns by an insulating material—which may be air in certain cases—and a difference of potential always exists between turns when the coil is conducting current, there is always a capacitive effect between the turns. This is shown in Fig. 3.7. This



**Fig. 3.7** Capacitive effects between the turns of a coil (distributed capacitance).

capacitive effect is known as the distributed capacitance of the coil, because it is distributed throughout the entire length of the coil.

Distributed capacitance causes a decrease in the effective inductance of a coil and an increase in its effective resistance. Hence, it is another undesirable effect, and measures are usually taken to decrease it as much as possible.



## Chapter 4

# Magnetic Circuits

### 4.1 INTRODUCTION

Today it is hardly possible to be concerned with electronics without at the same time being concerned with magnetism or electromagnetism in one form or another. Consider, for instance, the numerous applications of electromagnetics in the modern television receiver with its loudspeaker, centering magnet, purity and focus magnets, deflection yoke, and various types of transformers and chokes. In other branches of electronics such as computers or navigational instruments—to mention only two—the importance of magnetism and electromagnetism is well established. The electronics industry is dependent (at least in part) on the supporting roles played by companies that specialize in the manufacture of magnetic and electromagnetic materials and components. Figure 4.1 shows a comprehensive selection of various magnetic materials including magnets, cores, bobbins, and iron-powder cores.

### 4.2 MAGNETS, NATURAL AND ARTIFICIAL

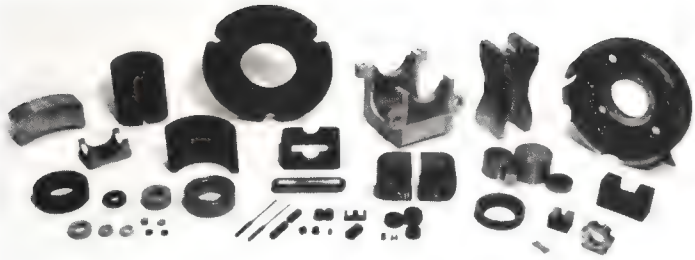
Natural magnets are found in the form of ore deposits known as *magnetite*, or more commonly, *lodestone*. This form of magnetic substance was known as early as 600 B.C. All other magnets are artificial. The first artificial magnets were made by contacting pieces of iron with magnetite. Even today, for some very minor applications, we may occasionally make a magnet by stroking a piece of iron or steel with any conveniently available magnet.

If a piece of soft iron is magnetized by induction from any magnetic source, the soft iron very quickly loses most of its induced magnetism after the magnetizing source is removed. However, if a piece of hardened steel or cast iron is magnetized, the induced magnetism will be retained for extremely long periods of time. Magnetized materials that very quickly lose most of their induced magnetism are called *temporary* magnets. Those that retain their magnetism are *permanent* magnets. Permanent magnets are made of wrought iron, hardened steel, or the various kinds of alloyed steels used for permanent magnets in loudspeakers and electric meters.

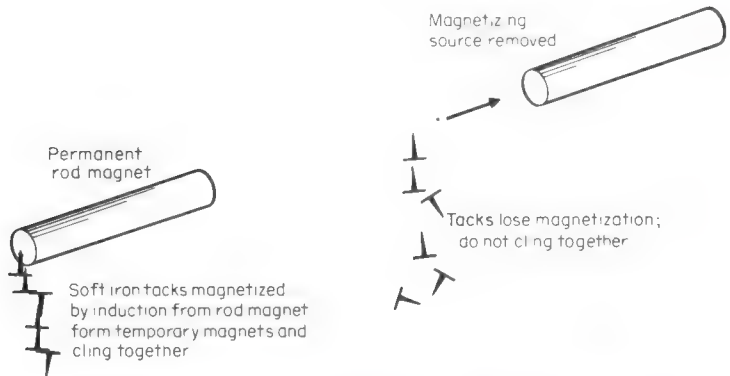
Figure 4.2 shows the difference between temporary and permanent magnets. Figure 4.3 illustrates a variety of permanent magnets which are used for making electric motors and other electric and electronic components.

### 4.3 MAGNETIC MATERIALS

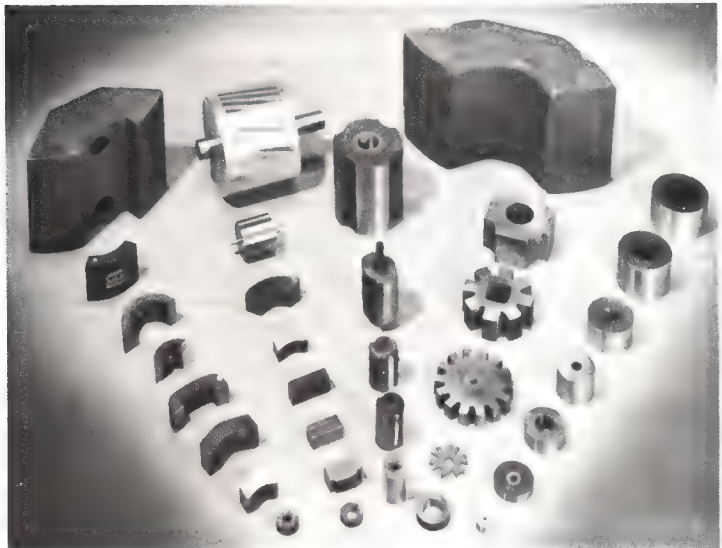
When we speak of magnetic materials, we usually mean substances that are very strongly attracted by magnets or magnetic fields. Substances of this type are said to be either *ferromagnetic* or *ferrimagnetic*. Ferromagnetic substances are the various kinds



**Fig. 4.1** A selection of various types of magnets, cores, bobbins, and iron-powder cores. (Courtesy The Arnold Engineering Company)



**Fig. 4.2** Comparison of permanent and temporary magnetism.

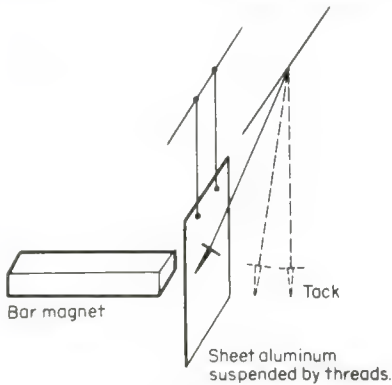


**Fig. 4.3** Permanent-magnet devices, used for motors and for other applications.

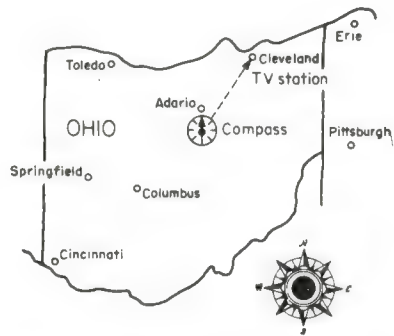
of iron and steel, including the powdered iron core materials that are used in some radio-frequency coils. Ferrimagnetic substances are the ferrites and other magnetic oxides that are used as core materials in coils operating at microwave frequencies, in high-frequency pulsing transformers such as the TV flyback transformer, and in the memory and switching units of high-speed computers.

Magnetic materials may also be classified as being either magnetically *soft* or magnetically *hard*. The magnetically soft materials are the kinds used in the cores of power and audio-frequency transformers and chokes. Magnetically hard materials are used for making permanent magnets.

Substances, such as aluminum, chromium, manganese, and air on which even very intense magnetic fields have only a very mild, barely detectable, attractive effect are called *paramagnetic*. Some substances, such as bismuth, antimony, copper, silver, and a few others on which very intense magnetic fields have a scarcely perceptible repellent effect, are said to be *diamagnetic*. Most substances, with the exception of those that are either ferromagnetic or ferrimagnetic, are commonly said to be *nonmagnetic*. Non-magnetic substances do allow magnetism to pass through them, but they never become magnetized to any noticeable degree. This is illustrated in Fig. 4.4.



**Fig. 4.4** Magnetic lines of force from the bar magnet pass through the nonmagnetic material (sheet aluminum) and attract the magnetic material (iron tack).



**Fig. 4.5** A compass and a map can be used for orienting a TV antenna.

#### 4.4 THE MAGNETIC COMPASS

If a piece of magnetite is suspended by a thread, it will align itself so that one part of it will always point in the general direction of the Earth's north magnetic pole. This was the very earliest form of compass, to be followed later by an artificial magnet suspended by a string. The modern magnetic compass takes the form of a magnetized steel needle balanced on a jeweled pivot and free to rotate horizontally. The end of the compass needle which points toward the north is called the *north seeking pole*. The other end, of course, points toward the south pole.

The compass, in conjunction with a map, is sometimes used to orient a TV antenna for reception from some distant station. The technique is illustrated in Fig. 4.5. The antenna being oriented is located at the point where the compass is lying on the map, and the desired station is in Cleveland. The dotted line shows the direction in which the antenna must be pointed to receive the signal directly from the station.

A compass may also be used in this manner to beam a radio signal toward some specific location.

#### 4.5 MAGNETIC POLES

When the north pole of one magnet is placed near to the north pole of another magnet, there is a repelling force between them. Likewise, there is a repelling force between two south magnetic poles, but a north pole and a south pole are strongly attracted. This

## 4-4 Magnetic Circuits

characteristic of magnetism is expressed in the law of magnetic attraction and repulsion which states:

**RULE:** *Magnetic poles that are alike will repel each other, while those that are unlike will attract each other.*

**COULOMB'S LAW** Coulomb's law, which defines the *amount* of attraction or repulsion, may be stated as follows:

**RULE:** *The force of attraction or repulsion between two magnetic poles is inversely proportional to the square of the distance between the poles and directly proportional to the product of the pole strengths.*

**THE UNIT MAGNETIC POLE** To use Coulomb's law in computations, we must have some method of precisely designating the strength of a particular magnetic pole in terms of some unit. This unit is the unit pole, derived from Coulomb's law and based on the concept that two magnetic poles of equal strength could be so selected that they would repel each other with a force of one dyne when the distance between the poles is exactly one centimeter. It is from this concept that the following definition of the unit magnetic pole is obtained:

The unit magnetic pole is of such strength that it will repel an exactly similar pole with a force of one dyne when the distance between the poles is one centimeter.

**MAGNETIC POLE COMPUTATIONS** From Coulomb's law we get the following equation for computing the force of attraction or repulsion between two magnetic poles:

$$F = \frac{M_1 M_2}{d^2} \quad (4.1)$$

where  $F$  = force between poles, dyn

$M_1$  = strength of first pole, unit poles

$M_2$  = strength of second pole, unit poles

$d$  = distance between poles, cm

**example 4.1** The force of repulsion between the north poles of two rod magnets is 100 dyn when the distance between them is 2 cm. What is the force of repulsion between the rod magnets if the distance between them is decreased to 1 cm?

**solution** In order to solve this type of problem, it is necessary to assume that the south magnetic poles of the rods are at a sufficient distance from the north poles so that their effects are negligible. In other words, the south pole of one of the rods is not noticeably attracted by the north pole of the other. Denoting the quantity to be determined by  $X$ , and using Coulomb's law, an inverse proportion can be written as

$$100 \cdot X = 1^2 \cdot 2^2$$

or

$$\frac{100}{X} = \frac{1^2}{2^2}$$

Solving for  $X$ , the result is

$$\frac{(100)(2^2)}{1^2} = \frac{(100)(4)}{1} = 400 \text{ dyn}$$

**example 4.2** The force of repulsion between the north poles of two bar magnets is 75 dyn when the distance between them is 10 cm. If the north pole of one of the bar magnets has a strength of 100 unit poles, what is the strength of the other bar magnet?

**solution** Again, it must be assumed that only the north poles of the magnets are exerting a measurable force, and the attraction or repulsion of the south poles can be ignored. Substituting into Eq. (4.1).

$$75 = \frac{100 \times M_2}{10^2}$$

and solving this for  $M_2$ , the result is

$$\frac{10^2 \times 75}{100} = \frac{100 \times 75}{100} = 75 \text{ unit poles}$$



## 4.6 THE EARTH'S MAGNETIC FIELD

From what is known about magnetic poles, it is evident that the only reason why a magnetic compass functions as it does is that the Earth itself is a huge magnet, with one of its poles in the north and the other just opposite, in the south. The Earth's magnetic pole that is located in the *geographical* north is actually a south *magnetic* pole. This is evident because it attracts the north pole of the magnetic compass.

Because the magnetic poles of the Earth are some distance from its geographic poles, the magnetic compass does not point precisely north from all points of the Earth's surface. In the eastern parts of the United States the compass will point west of geographic or true north; and in the western parts of the United States, it will point east of true north. The angle by which a compass points away from geographic or true north is known as *magnetic declination*. Tables and charts showing the magnetic declination in degrees at various points of the Earth's surface may be found in handbooks on surveying and navigation.

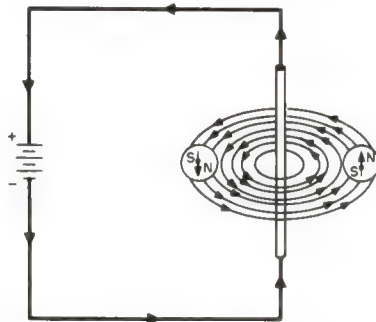
Some knowledge of the Earth's magnetic field and the compass is very often of importance to those working with or on electronic navigational equipment, radar or other locating equipment, missile tracking, and radio astronomy.

**example 4.3** In making tests on a gyrocompass, it is necessary to orient it so that it points to geographic or true north. A precise magnetic compass is available. What procedure should be followed?

**solution** The gyrocompass should first be oriented to point to magnetic north, using the magnetic compass. The necessary correction should then be made by consulting a table of magnetic declinations.

## 4.7 ELECTROMAGNETISM

If a compass is brought near a conductor carrying current, as shown in Fig. 4.6, the compass needle will point in the direction shown. If the current through the conductor



**Fig. 4.6** The magnetic field encircling a current-carrying conductor. (Conventional current shown by arrows in circuit.)

is reversed, the compass needle will turn and point in the opposite direction. From this it may be seen that a conductor carrying electric current is surrounded by a magnetic field, and this field exhibits bipolar effects. In other words, the field has direction, just as the magnetic field of a permanent magnet.

The magnetic field is said to come out of the magnet at its north pole and enter at its south pole. Actually the field lines do not move. Instead, their direction is defined as the direction that a unit north pole would move if placed in the field. There is no such thing in real life as a unit north pole, but it is a useful imaginary concept for defining the direction of magnetic fields.

To summarize, the magnetic field around a current-carrying wire has a north-to-south direction just as the field around a permanent magnet has direction. This direction is



defined as the direction that a unit north pole will move, and is indicated by the direction a compass points when placed in the field.

**THE RIGHT-HAND RULE AND THE LEFT-HAND RULE FOR MAGNETIC FIELDS AROUND WIRES** To determine the direction of the magnetic field surrounding any current-carrying conductor, use the following rule which is illustrated in Fig. 4.7.

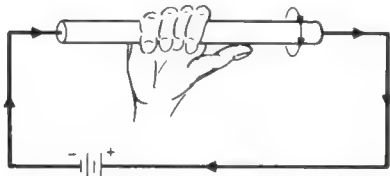
**RULE:** *Grasp the wire in the right hand, and let the thumb point in the direction in which conventional current (that is, current from positive to negative) is flowing through the wire. The fingers will then point in the same direction as the direction of the magnetic field, or lines of force, that encircle the wire.*

If electron current is assumed, the left-hand rule is used.

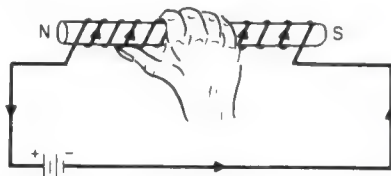
**RULE:** *Grasp the wire with the left hand so that the thumb points in the direction of electron current flow and the fingers will encircle the wire in the direction of the magnetic field.*

#### 4.8 THE RIGHT-HAND RULE AND THE LEFT-HAND RULE FOR COILS

Any coil or solenoid that carries current exhibits magnetic effects, and magnetic effects are always of a bipolar nature. Any current-carrying coil must always have both a north and a south pole. To determine the magnetic polarity of a current-carrying coil, use the right-hand rule for coils (see Fig. 4.8):



**Fig. 4.7** The right-hand rule for determining the direction of the magnetic field around a current-carrying conductor. (Conventional current shown.)

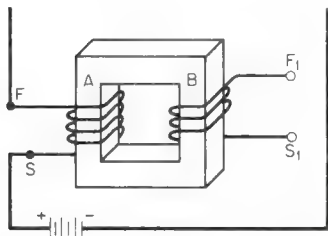


**Fig. 4.8** The right-hand rule for determining the magnetic polarity of a coil.

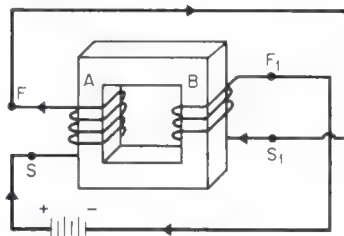
**RULE:** *Grasp the coil in the right hand, and let the fingers point in the direction of the conventional current around the coil; the thumb will then point in the direction of the north pole of the coil.*

If the coil is grasped with the left hand so that the fingers encircle the coil in the same direction that *electron* current is flowing, then the thumb of the left hand will point in the direction of the north pole.

**example 4.4** Two identical coils, A and B, are wound on an iron core of rectangular cross section, as shown in Fig. 4.9. The coils are wound in opposite directions, and the start of coil A is connected to the positive terminal of a battery. How should the other terminals of the two coils be connected together in series and to the negative terminal of the battery so that their magnetic fields will aid each other?



**Fig. 4.9** Connect the coils to the battery so that their magnetic fields combine.



**Fig. 4.10** Coils A and B are connected so that their fields are aiding.

**solution** Terminal  $F$  of coil  $A$  should go to terminal  $S_1$  of coil  $B$ , and  $F_1$  of coil  $B$  should be connected to the negative terminal of the battery, as shown in Fig. 4.10. Check this illustration with the right-hand (or left-hand) rule to make sure that the fields of the coils are aiding.

**example 4.5** What would be the effect if the two coils,  $A$  and  $B$ , were to be connected in parallel and to a battery, as in Fig. 4.11?

**solution** In this case also, the magnetic fields of the two coils would aid each other. Application of either the right-hand rule or the left-hand rule for coils will prove this.

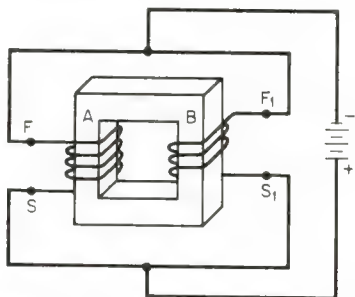


Fig. 4.11 Coils  $A$  and  $B$  are connected in parallel. Do the magnetic fields of the coils aid or oppose?

## 4.9 UNITS OF MEASUREMENT IN MAGNETICS

The basic units of measurement in electricity are amperes, volts, and ohms. It would be convenient if there were also only three basic units of magnetism, but unfortunately such is not the case. Instead, there are several different systems of measurement, and there is no universally used system. (The mks system has been suggested as a logical universal system of measurement in magnetism, but at the time of this writing the literature from American manufacturers does not define their products in terms of the mks system exclusively.)

One way to understand the myriad of units for magnetic measurements would be to describe all the units in a particular system. For example, the method of measuring flux, flux density, magnetomotive force, etc., in the cgs system could be described carefully, and then a conversion table could be included at the end of the discussion for changing from cgs to any other system.

Another method—the one used in this book—is to describe each term and its unit of measurement in the system that describes that particular term most simply. Thus, we will define the unit of *flux* in terms of the cgs system, and the unit of magnetomotive force in terms of the mks system. A table will then be included for converting from one system to another.

**MAGNETIC FLUX** Magnetic fields exist in lines of force, called flux. In the cgs system, each individual line of flux is called a *maxwell*. If there are three lines of flux, the amount of flux is said to be three maxwells in the cgs system, whereas in the British system it is simply called three *lines*. The greater the number of lines of flux, the stronger the magnetic field. The lines are imaginary, and the concept of lines of flux probably comes from the popular experiment in which iron filings are sprinkled on a paper that is placed over a magnet. The iron filings tend to arrange themselves in a group of lines, as shown in Fig. 4.12. These are actually lines of equal magnetic intensity.

**UNIT POLES** All magnets contain a north and a south pole. Some magnets can be made to have more than one north and south pole, but there is always a south pole for each north pole. For the purpose of discussion, however, let us suppose that it is possible to make an extremely small north pole with no associated south pole. This very small imaginary north pole is called a unit north pole in the cgs system. This unit north

pole is helpful in describing some of the terms used in magnetism, and some of the units of measurement. However, it must always be remembered that it does not actually exist anywhere in the world physically.

How small should this unit north pole be? It is not enough to just say small, or very small; it is necessary to say exactly how small if it is going to be useful for making measurements, and for describing units of measurement. Figure 4.13 shows how the unit north pole is defined. If two unit north poles are placed in a vacuum one centimeter apart, they will *repel* each other with a force of one dyne. This defines, then, how small a unit north pole is.

If a unit north pole was placed on the piece of paper in Fig. 4.12, it would move away from the north pole of the magnet and toward the south pole. Its path will be along one of the flux lines shown in the illustration. For this reason, you will often see arrows along the lines of flux. These arrows are supposed to show the direction of a unit north

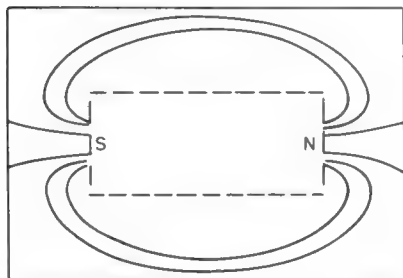


Fig. 4.12 Arrangement of iron filings on a piece of paper covering a bar magnet.

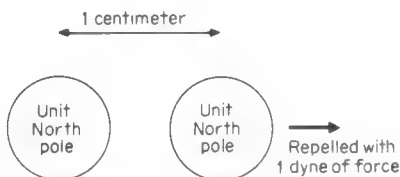


Fig. 4.13 Definition of unit north pole.

pole along the flux line, but often give the mistaken idea that the flux line itself is supposed to be moving. The direction of a line of flux is defined as the direction that a unit north pole will move if placed on the flux line. One line of flux, or one maxwell, will act upon a unit north pole with a force of one dyne. Thus, the unit north pole is helpful in defining the magnetic strength of a line of flux.

**FLUX DENSITY** If one line of flux passes perpendicularly through a square centimeter of this page, then there is a certain amount of magnetic field strength present. If two lines of flux pass through that same square centimeter, then the field strength will be twice as great as when one flux was present. If three lines of flux pass through that square centimeter, the field strength will be three times as great, and so on. It should be easy to visualize that three lines of flux passing through one square centimeter will produce a greater concentration of magnetic field strength than would be produced if the two lines of flux pass through one square centimeter. An important way of defining the strength of a magnetic field, then, is in terms of the number of flux lines that pass through a square centimeter of area. Obviously, a very strong magnet will cause more flux lines to pass through a square centimeter of area than a weak magnet will. This is illustrated in Fig. 4.14.

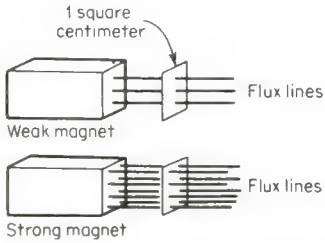
The number of lines of flux that pass through one square centimeter of area is called the flux density. In the cgs system, the unit of flux density is the *gauss*. When one maxwell (that is, one line of flux) passes through one square centimeter of cross-sectional area, the flux density is said to be one gauss. Figure 4.15 illustrates one gauss of flux density. The gauss is a unit of measurement in the cgs system.

**MAGNETOMOTIVE FORCE** Flux lines can be established in a number of ways. One of the easiest ways is to use an electric current. For every electric current there is an associated magnetic field. The force that causes a magnetic flux to be established is called the magnetomotive force. It can be compared to the voltage in electricity which is sometimes considered to be the "force" that produces an electric current. (Actually, of course, voltage is a measure of work, *not* a unit of force.)

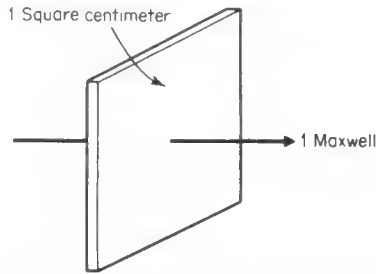
The easiest unit of magnetomotive force to understand is the *ampere-turn*. If one ampere of current flows through one turn of wire, then the magnetomotive force is said to be one ampere-turn.

Figure 4.16 illustrates the ampere-turn. The magnetomotive force is always equal to the number of amperes of current multiplied by the number of turns of wire. The ampere-turn is a mks unit of measurement. (In the cgs system, a gilbert is used as the unit of mmf. One gilbert =  $1.257 \times$  ampere-turns.)

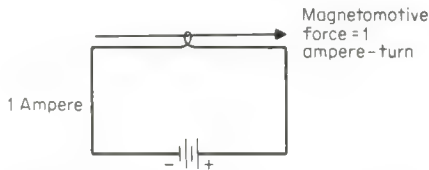
**RELUCTANCE** When a magnetomotive force is used to establish flux in a material, there is always some opposition to the flux. This opposition is called *reluctance*. There



**Fig. 4.14** The strong magnet will cause more lines of flux per square centimeter than the weak magnet.



**Fig. 4.15** Illustrating one gauss of flux density.



**Fig. 4.16** The ampere-turn is a unit of magnetomotive force.

are no *English* units for measuring reluctance. The term *rel* has been proposed but is not in common usage. In the cgs system reluctance is measured in *gilberts per maxwell*, but there is no common name for the unit of reluctance, and there is no name for it in the mks system.

In magnetic circuits, it is sometimes more convenient to know the ease with which flux lines can be established in a material, rather than the opposition that the material offers to establishment of flux. The ease with which flux lines can be established in a material is called the *permeance* of that material, and it is the reciprocal of reluctance.

**COMPARISON OF ELECTRIC AND MAGNETIC CIRCUITS** There is a similarity between electric and magnetic units. Table 4.1 lists the more important electrical terms and the comparable magnetic terms.

**TABLE 4.1 Comparison of Electric and Magnetic Terms**

Electric units	Magnetic units
Electric current $I$	Magnetic flux $\phi$
Electromotive force $E$	Magnetomotive force, mmf
Resistance $R$	Reluctance $R$
Conductance $G$	Permeance $P$

There are certain things about electric circuits that are quite different from magnetic circuits, and should always be kept in mind. The more important differences will now be discussed.

In an electric current, all the current is normally confined to the conductors carrying it. In a magnetic circuit, some of the flux lines actually leave the circuit along the route, so that it is somewhat more difficult to calculate the amount of flux at a certain point in comparison to calculating current in a circuit.

The resistance of a wire or resistor is a fixed value for a particular circuit over the range of currents for which it is intended to be used. In a magnetic circuit, the amount of reluctance is dependent upon the amount of flux already present. The greater the amount of flux already present in a piece of iron, the greater the reluctance becomes—that is, the greater the opposition to a further increase in flux.

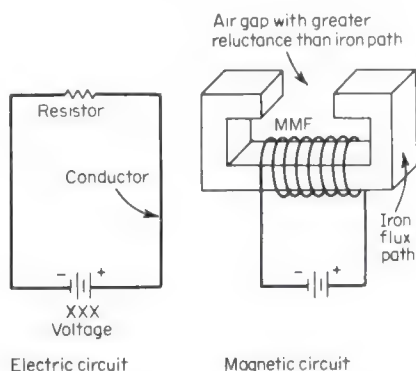


Fig. 4.17 Comparison of an electric and a magnetic circuit.

In electric circuits, there are insulating materials that offer great opposition to current flow. In magnetic circuits there are no materials that can prevent flux from being established. In order to prevent magnetic flux lines from entering a particular area, the only course that can be taken is to employ a very low reluctance path around that area. Nearly all the flux lines will follow the low-reluctance path in such cases, and therefore they will not enter the region being shielded.

Figure 4.17 compares an electric and a magnetic circuit. A battery supplies the power to the electric circuit. In the magnetic circuit the coil of current-carrying wire provides the magnetomotive force that establishes flux in the circuit.

Electric current flows through the conductor in the electric circuit. In the magnetic circuit a low-reluctance iron provides a path for the flux. As mentioned before, some of the flux may leave this low-reluctance path and return to the source of mmf by another path. If the circuit is properly designed, the flux leakage can be kept to a minimum.

A resistor in the electric circuit offers opposition to the flow of current. In the magnetic circuit, the air gap offers a greater opposition to establishment of flux than the iron path. Therefore, the air gap may be thought of as being the resistor in the magnetic circuit.

Table 4.2 lists some of the units of magnetism, and conversion factors between units.

**MAGNETIZING FORCE** One ampere of current flowing through one turn of wire will give a magnetomotive force of one ampere-turn. (Mmf in ampere-turns = number of amperes  $\times$  number of turns of wire.) Likewise, one ampere of current flowing through ten turns of wire will give a magnetomotive force of ten ampere-turns. Now, suppose you wind the ten turns very closely together so that each turn is tight against the adjacent turns. The magnetomotive force will then be concentrated along a short length of the coil form. On the other hand, if you spread the turns out so that they are evenly distributed along a length of one foot, then the magnetomotive force will be



TABLE 4.2 Magnetic Units

Unit	cgs	mks	British
Flux	Maxwells	Webers	Lines
Magnetomotive force	Gilberts	Ampere-turns	Ampere-turns
Magnetizing force	Oersteds	Ampere-turns per meter	Ampere-turns per foot
Reluctance	Gilberts per maxwell	Ampere-turns per weber	No unit (rel suggested)
Flux density (also called magnetic induction)	Maxwells per square centimeter <sup>2</sup>	Webers per square meter	Lines per square inch

## CONVERSION FACTORS

Inches $\times 2.54$ = centimeters	Gilberts $\times 10/4\pi$ = ampere-turns
Maxwells $\times 10^{-8}$ = webers	Oersteds $\times 10^3/\pi$ = ampere-turns per meter
Gauss (maxwells per square centimeter) $\times 10^{-4}$ = webers per square meter	
1 gamma = $10^{-5}$ gauss	

\* One maxwell per square centimeter is called a gauss; so flux density in the cgs system can be given in either the number of maxwells per square centimeter, or gauss.

distributed along a path one foot long. In both cases *the magnetomotive force* is ten ampere-turns, but in one the mmf is concentrated in a short length, whereas in the other it is spread out along a length of one foot.

It is important to know the amount of magnetomotive force, but it is also essential to know how it is distributed. The distribution of mmf along a length is called the *magnetic gradient*. It is also known as the *magnetizing force*. In the above example, ten turns per inch will produce a greater magnetizing force than ten turns per foot. In the mks system, magnetizing force is measured in ampere-turns per meter, while in the British system either ampere-turns per foot or ampere-turns per inch may be used. In the cgs system, the oersted is used. One gilbert per centimeter gives a magnetizing force of one oersted.

**RELATIVE PERMEABILITY** Relative permeability, which is often simply called permeability, is the ratio of the flux density in a material for a given magnetizing force to the flux density that would be produced in a vacuum for the same amount of magnetizing force. Mathematically, the relative permeability in the cgs system is expressed by the equation

$$\mu = \frac{B}{H} \quad (4.2)$$

where  $\mu$  = relative permeability of the material

$B$  = flux density

$H$  = magnetizing force

There are no units for relative permeability. Relative permeability depends on the type of material. In other words, every different type of material will have a different value of relative permeability. For example, the relative permeability of air, like that of a vacuum, is one. For all paramagnetic materials, the relative permeability is slightly greater than one. The relative permeability of diamagnetic materials is less than one, while the relative permeability of ferromagnetic materials is much greater than one (and also much greater than the relative permeability of paramagnetic materials).

A relative permeability greater than one simply means that it is easier to establish flux lines in that material than it is to establish them in a vacuum. A material having a relative permeability less than one indicates that it is more difficult to establish flux lines in that material than it is to establish them in a vacuum.





## Chapter 5

# Transformers

### 5.1 INTRODUCTION

The function of every type of transformer is to transform or change electricity in some manner. This is true whether the transformer is only a tiny sub-ouncer unit or is one of the huge devices used at the substations and terminal points of high-voltage electric-power distribution systems. For example, the tiny output transformer of a transistor pocket-portable radio transforms the current in the output stage to a form that is suitable for actuating the loudspeaker. As another example, the large transformer at the power distribution substation transforms (steps down) the high voltage of the power lines so that it may safely be used in homes and industry.

Since the iron-cored transformer and choke came into use much earlier than other kinds of transformers and chokes, it seems only fitting to take up the study of such transformers and chokes (see Fig. 5.1) before taking up the study of other types. Figure 5.1a shows a variety of ferrite-core coils. An iron-core transformer is shown in Fig. 5.1b.

### 5.2 THE BASIC TRANSFORMER

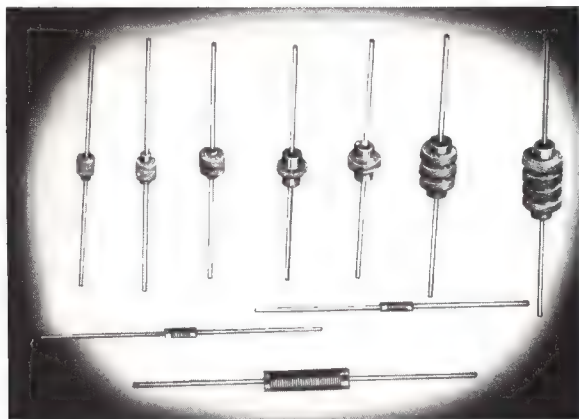
The basic transformer is shown in Fig. 5.2. It consists of two coils wound upon the same iron core which forms a closed magnetic circuit. The flux path is shown by dotted lines around the core. The transformer will not work on pure direct current, but instead must be supplied with power from a source of alternating current or from a source of pulsating direct current.

The winding to which power is supplied is called the *primary*, and the winding from which power is taken is the *secondary*. Any change in the primary current results in a change in flux ( $\phi$ ). In accordance with Faraday's law, a change in the flux linking the secondary winding will induce a voltage across the secondary. If a path for current flow exists across the secondary winding, then in accordance with Lenz's law the direction of current flow will be such that its magnetic field opposes the changing flux that produced it.

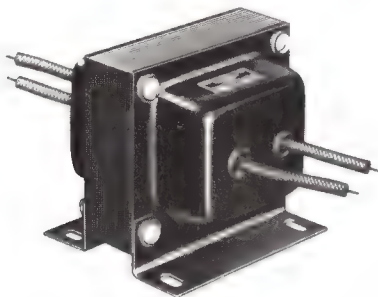
If the secondary voltage is the same as the primary voltage, the transformer is called a *one-to-one transformer*; if the secondary voltage is greater than the primary, it is a *step-up transformer*, and when the secondary voltage is less than the primary, it is a *step-down transformer*. Power supplied to the primary at a specific frequency produces power at the same frequency in the secondary.

### 5.3 MUTUAL INDUCTION

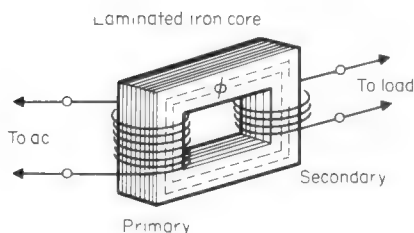
If the transformer primary of Fig. 5.2 is connected to a source of alternating current, as in Fig. 5.3, the instantaneous values of the primary voltage and current and the resulting



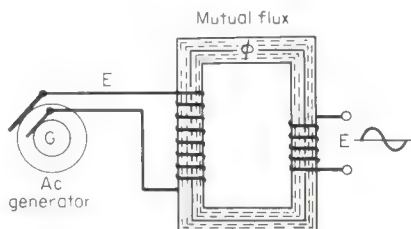
**Fig. 5.1a** A variety of ferrite-core coils and chokes. (Courtesy James Millen Manufacturing Company, Inc.)



**Fig. 5.1b** Iron-core transformer with end bell construction. (Courtesy Acme Electric Corporation.)



**Fig. 5.2** Illustrating the basic transformer. Power is transferred from the primary to the secondary winding by means of the continuously changing flux ( $\phi$ ) in the core.



**Fig. 5.3** When the transformer is connected to a source of alternating current, the continuously changing magnetic flux in the core induces a voltage in the secondary winding and at the same time also induces a back emf (or counter-emf) in the primary winding.

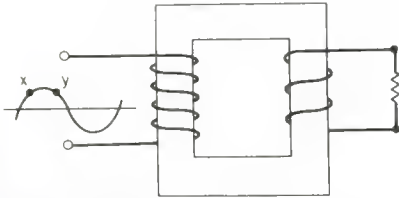
magnetic field will all be continuously changing and inducing a back voltage in the primary. The instantaneous value of the back voltage will be very nearly equal to the instantaneous value of the applied voltage, but it will always be just a bit lower than the applied voltage. If the applied and back voltages were exactly equal to each other, it would be impossible to get the current flowing in the circuit.

Since the core of the transformer forms a closed magnetic circuit, the primary and secondary windings are interlinked through the mutual magnetic flux in the core. The changing flux in the core induces a voltage in the secondary at the same time that it is inducing a back voltage in the primary, and the combined induction effect is called mutual induction.

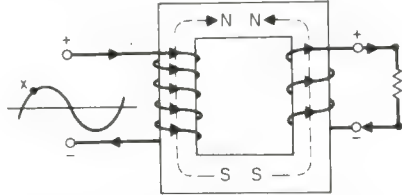
The secondary induced voltage may be the same as the primary voltage, or it may be higher or lower than the primary voltage, depending on the ratio of the number of turns of wire between primary and secondary.

When the secondary is connected to a load, current flows from the secondary into the load just as it would from any other voltage source. According to Lenz's law, the current flow in the secondary must always be in such a direction that it will set up an opposition to any *change* in the flux of primary current. For example, if the secondary has a load connected to it, the load current in the secondary winding creates a magnetic field, and as this field changes, it induces a voltage in the primary in opposition to the applied voltage.

**example 5.1** Figure 5.4 represents a transformer supplying power to a load connected to its secondary. A waveform of the current is also shown. Draw a diagram showing terminal polarities of both primary and secondary windings and the direction of current flow in each



**Fig. 5.4** Transformer with sine-wave current input supplying a load.



**Fig. 5.5** Terminal polarities and the directions of the magnetic fields produced by currents in the primary and secondary windings at the instant when the primary current is rising and is at point *x* on the current curve.

winding at the instant when the primary current is rising from zero to its maximum positive value and has reached the point *x* on the current curve. Also show the directions of the magnetic fields developed by each winding at this instant. Use the right-hand rule for coils and Lenz's law.

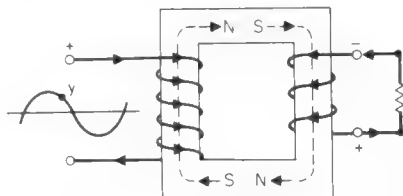
**solution** Figure 5-5 shows that as the current is rising from zero to its maximum positive value, the upper terminal of the primary will be positive. Therefore, current flow through the primary winding will be in the direction shown by the arrows, and by the right-hand rule for coils, the direction of the resulting magnetic field will be clockwise in the core. Since the field is building up, the direction of the magnetic field developed in the secondary must, according to Lenz's law, oppose the building up of the field developed by the primary; hence, the magnetic field developed by the secondary must be in a counterclockwise direction in the core. The right-hand rule for coils will show that the magnetic field developed by the secondary will be in a counterclockwise direction in the core when current flows through the coil windings in the direction indicated by the arrows. The upper terminal of the secondary will then be positive.

**example 5.2** Draw a diagram showing conditions in the transformer of Fig. 5.4 at the instant when the primary current is falling from its maximum positive value toward zero and has reached the point *y* on the current curve.

**solution** The upper terminal of the primary winding is still positive and the primary current will be in the direction indicated. Since the current is now decreasing, its associated magnetic field will be collapsing. According to Lenz's law, the magnetic field developed in the secondary must oppose any *decrease* of the primary field. Therefore, the magnetic field of the secondary winding at this instant will be in the same direction in the core as the primary field, or clockwise, as shown in Fig. 5.6. Application of the right-hand rule for coils will

## 5-4 Transformers

then show that current flow in the secondary will be in the direction indicated by the arrows, and the lower terminal of the secondary winding will be positive.



**Fig. 5.6** Showing terminal polarities and the directions of the magnetic fields produced by currents in the primary and secondary windings of a transformer at the instant when the current is falling and has reached point *y* on the curve of current.

### 5.4 TRANSFORMER VOLTAGE RELATIONSHIPS

Since the mutual flux in the core of any transformer interlinks the turns of the primary and secondary windings, any change in this flux that induces a certain voltage in *each* turn of the primary winding must also induce the same voltage in *each* turn of the secondary winding. Then, if the secondary winding has only half as many turns as the primary, total secondary voltage will be only half the induced or impressed primary voltage. If the secondary has twice as many turns as the primary, the total induced secondary voltage will be twice as great as the impressed primary voltage.

It may be concluded, then, that the total voltage induced in each winding is always proportional to the number of turns in the winding:

$$\frac{E_p}{E_s} = \frac{N_p}{N_s} \quad (5.1)$$

where  $E_p$  = impressed primary voltage

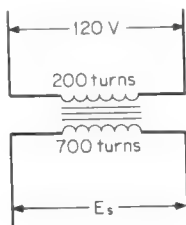
$E_s$  = secondary voltage

$N_p$  = number of primary turns

$N_s$  = number of secondary turns

When there is no load connected to the secondary, the induced primary voltage has practically the same value as the impressed voltage; and even when the transformer is supplying a load, if the transformer is properly designed, there will be only a very slight difference between the impressed and induced primary voltages. In an efficient, well-designed transformer the voltage drops due to resistance and reactance will be very low. Therefore, the ratio  $E_p/E_s$  or  $N_p/N_s$  may be taken as the voltage ratio of the transformer.

**example 5.3** What is the secondary voltage of the transformer shown in Fig. 5.7?



**Fig. 5.7** What is the voltage across the secondary?

**solution** Substituting in Eq. (5.1),

$$\begin{aligned}\frac{E_p}{E_s} &= \frac{N_p}{N_s} \\ \frac{120}{E_s} &= \frac{200}{700}\end{aligned}$$

Solving for the unknown,

$$E_s = \frac{120 \times 700}{200} = 420 \text{ V} \quad \text{Answer}$$

**example 5.4** The primary winding of a transformer contains 100 turns. How many turns must the secondary contain to step down a voltage of 120 to 12.6 V?

**solution** By substitution in Eq. (5.1),

$$\frac{120}{12.6} = \frac{100}{N_s}$$

Solving for  $N_s$ ,

$$N_s = \frac{100 \times 12.6}{120} = 10.5 \text{ turns}$$

## 5.5 TRANSFORMER CURRENT RELATIONSHIPS

The modern transformer is a highly efficient device. In fact, some large iron-core transformers have an efficiency exceeding 99 percent. This means that almost as much power may be taken out of the transformer as is put into it. Therefore, for practical purposes we may assume that the transformer has an efficiency of 100 percent. If the efficiency is 100 percent, primary and secondary power factors are equal. Power input to the transformer is given by  $E_p I_p \times (\text{power factor})$ , and the output power is  $E_s I_s \times (\text{power factor})$ . Assuming that the input and powers are equal,

$$E_p I_p \times (\text{power factor}) = E_s I_s \times (\text{power factor})$$

Since power factor in the first member has the same value in the second member of the equation, it cancels out, leaving  $E_p I_p = E_s I_s$  or

$$\frac{E_p}{E_s} = \frac{I_s}{I_p} \quad (5.2)$$

where  $E_p$  = primary impressed voltage

$I_s$  = secondary current

$E_s$  = secondary voltage

$I_p$  = primary current

In a step-up transformer, if the secondary voltage is twice the primary voltage, secondary current is only one half the primary current.

Since

$$\frac{E_p}{E_s} = \frac{N_p}{N_s}$$

it follows that

$$\frac{N_p}{N_s} = \frac{I_s}{I_p} \quad (5.3)$$

Equation (5.3) shows that the currents in the primary and secondary windings are inversely proportional to the corresponding turns.

**example 5.5** A transformer supplies 2.4 A at 6.3 V to a secondary load. What is the primary current if the primary is connected to a 120-V ac source?

**solution** By substitution in Eq. (5.2),

$$\begin{aligned}\frac{E_p}{E_s} &= \frac{I_s}{I_p} \\ \frac{120}{6.3} &= \frac{2.4}{I_p}\end{aligned}$$

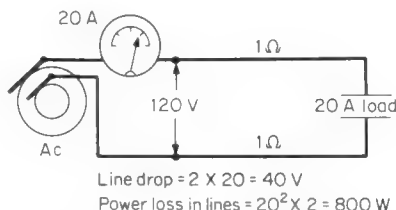
## 5-6 Transformers

Solving for  $I_p$ ,

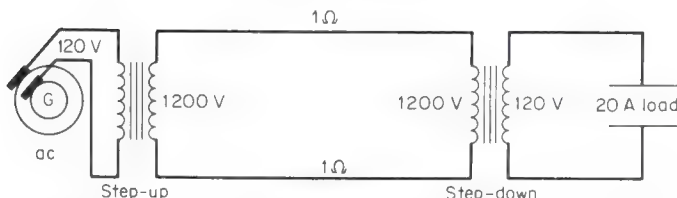
$$I_p = \frac{6.3 \times 2.4}{120} \\ = 0.126 \text{ or } 126 \text{ mA}$$

*Answer*

**example 5.6** A 120-V generator supplies a current of 20 A to a load some distance from the source. The total resistance of the transmission lines from the generator to the load is  $2 \Omega$ . The line drop, in this case, is therefore  $2 \times 20$ , or 40 V, and the voltage across the load is only  $120 - 40$ , or 80 V. The power loss in the lines is  $20^2 \times 2$ , or 800 W. These conditions are shown in Fig. 5.8.



**Fig. 5.8** Resistance of line wires between generator and load causes large power loss.



**Fig. 5.9** Stepping up the line voltage at the generator results in reduction of line loss to insignificant value. Line current = 2 A; line drop =  $2 \times 2 = 4$  V; power loss in lines =  $2^2 \times 2 = 8$  W.

Using the same transmission lines, what will be the line drop and the line loss if a transformer is used to step up the voltage at the generator to 1200 V, and another transformer at the load is then used to step down the line voltage from 1200 to 120 V, as shown in Fig. 5.9.

**solution** Since the load takes 20 A at 120 V, and the primary of the load transformer operates at 1200 V,

$$\frac{E_p}{E_s} = \frac{I_s}{I_p} \quad (5.2)$$

$$\frac{120}{1200} = \frac{I_s}{20}$$

$$I_s = \frac{20 \times 120}{1200} = 2 \text{ A} \quad \text{Answer}$$

The line drop is found by Ohm's law:

$$E = IR = 2 \times 2 = 4 \text{ V} \quad \text{Answer}$$

The power loss is

$$P = I^2 R = (2)^2 \times 2 = 8 \text{ W} \quad \text{Answer}$$

## 5.6 THE FUNDAMENTAL EQUATION OF THE TRANSFORMER

The average of any varying quantity is the average of all the instantaneous values that the variable will assume in changing from one value to another.

If the voltage impressed on the primary of a transformer has a sine waveform, as shown in Fig. 5.10, this will produce similar sine waves of both current and flux. (This assumes that the flux in the magnetic circuit is linearly related to the current in the

winding. Although this is usually not true, it is a reasonable assumption when the flux does not approach the saturation point.) The average value of the induced voltage in either the primary or secondary of a transformer is given by

$$E_A = \frac{N\phi_m}{10^8 t} \quad (5.4)$$

where  $E_A$  = average induced voltage

$N$  = number of turns

$\phi_m$  = maximum flux

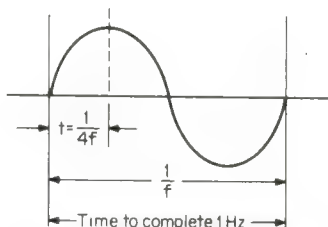
$t$  = time in seconds for the flux to rise from zero to its maximum value

Figure 5.10 shows that a sine wave rises from zero to maximum in one-fourth the time it takes to complete one cycle. Therefore, for a sine wave of frequency  $f$

$$t = \frac{1}{4f}$$

Equation (5.4) may be written as

$$E_A = \frac{N\phi_m}{10^8 (1/4f)} = \frac{4fN\phi_m}{10^8} \quad (5.5)$$



**Fig. 5.10** A voltage or current having a sine-wave form rises from zero to maximum in one-fourth of the time it takes to complete 1 Hz.

To change Eq. (5.5) to a form that gives effective or root-mean-square voltage  $E$ , instead of average values, we multiply by

$$\frac{0.707}{0.636}$$

Thus,

$$\begin{aligned} E &= \frac{0.707}{0.636} \times \frac{4fN\phi_m}{10^8} \\ &= \frac{1.11 \times 4fN\phi_m}{10^8} \\ &= \frac{4.44fN\phi_m}{10^8} \end{aligned} \quad (5.6)$$

These equations show that the induced voltage is directly proportional to the maximum flux in the core, the frequency of the applied voltage, and the number of turns in the winding.

The permissible flux in a magnetic circuit is usually expressed in terms of the flux density  $B_m$ .

$$B_m = \frac{\phi_m}{A}$$

where  $B$  = lines per square inch if the area  $A$  is in square inches

$B_m$  = gaussses if the area is in square centimeters



## 5-8 Transformers

Then Eq. (5.6) may be written as

$$E = \frac{4.44fN_p B_m A}{10^8} \quad (5.7)$$

Equation (5.7) is the fundamental equation of the transformer. In the transformer primary, the effective value of the induced voltage,

$$E_p = \frac{4.44fN_p B_m A}{10^8} \quad (5.8)$$

In the transformer secondary, the effective value of induced voltage,

$$E_s = \frac{4.44fN_s B_m A}{10^8} \quad (5.9)$$

Effective values are of interest since transformer ratings are in these values, and all meters, unless marked otherwise, indicate effective values.

**example 5.7** The primary winding of a transformer contains 300 turns. Determine (a) at what maximum flux ( $\phi$ ) must the core operate if the impressed primary voltage is 120 V at 60 Hz? (b) How many turns must the secondary winding have to develop 12.6 V?

**solution** (a) The maximum flux  $\phi = B \times A$ ; by substitution in Eq. (5.8),

$$E_p = \frac{4.44 \times 60 \times 300 \times \phi_m}{10^8} = 120$$

Solving for  $\phi_m$ , we get

$$\phi_m = \frac{120 \times 10^8}{4.44 \times 60 \times 300} = 150 \text{ } 150 \text{ Mx} \quad \text{Answer}$$

(b) By substitution in Eq. (5.9),

$$E_s = 12.6 \frac{4.44 \times 60 \times N_s \times 150 \text{ } 150}{10^8}$$

Solving for  $N_s$ , the result is

$$\frac{12.6 \times 10^8}{4.44 \times 60 \times 150 \text{ } 150} = 31.5 \text{ turns} \quad \text{Answer}$$

**example 5.8** A transformer primary winding has 450 turns, and its secondary has 2700. The cross-sectional area of the core is 1 in<sup>2</sup>, and maximum flux density  $B_m$  in the core is 100 000 lines in<sup>2</sup>. Line frequency is 60 Hz. Compute: (a) rated primary voltage  $E_p$ , (b) rated secondary voltage  $E_s$ , (c) effective voltage induced in each primary turn, (d) effective voltage induced in each secondary turn.

**solution** (a) Substituting in Eq. (5.8),

$$E_p = \frac{4.44 \times 60 \times 450 \times 100 \text{ } 000 \times 1}{10^8} = 119.88$$

(b) Substituting in Eq. (5.9),

$$E_s = \frac{4.44 \times 60 \times 2700 \times 100 \text{ } 000 \times 1}{10^8} = 719.28$$

$$(c) \frac{119.88}{450} = 0.266 \text{ V}$$

$$(d) \frac{719.28}{2700} = 0.266 \text{ V}$$

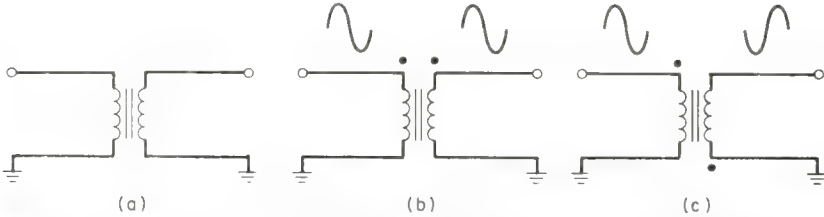
## 5.7 DISCUSSION OF NO-LOAD CONDITIONS

If the secondary winding of a transformer is left open-circuited, the primary current is very small, since the primary current is then opposed not only by the resistance of the winding but also by its inductive reactance which has a *choking* effect on the primary current. The no-load current produces the magnetic flux, and at the same time supplies the eddy current and hysteresis losses in the core. The no-load current, therefore, consists of two components: the magnetizing current component and a power-loss com-

ponent The magnetizing current component lags the applied primary voltage by  $90^\circ$ , but is always in phase with the flux. The core loss, or power component, is always in phase with the applied voltage.

Since the magnetizing current component lags the applied primary voltage by  $90^\circ$ , it must lead the induced primary voltage by  $90^\circ$ , because the primary applied and induced voltages are  $180^\circ$  out of phase with each other.

According to Lenz's law, the magnetizing current must lead the induced secondary voltage by  $90^\circ$ , since as the magnetizing current and flux decrease from their maximum positive values and pass through zero in a negative direction, both current and flux



**Fig. 5.11** Use of dot notation to indicate phase of input and output signals: (a) The transformer symbol does not indicate the phase of input and output signals. (b) The dots indicate leads having the same phase. (c) Connection of a transformer for phase reversal.

will be changing at their greatest rates at the exact instant when they are passing through their zero values. Therefore, the greatest rate of change of flux in the core occurs at this very instant, and the induced voltages in primary and secondary windings will both reach their maximum values at this instant. This should not be misconstrued to mean that the voltage across the secondary winding must always be  $180^\circ$  out of phase with the voltage across the primary.

Figure 5.11 shows how the phase of the secondary voltage is related to the phase of the primary voltage. The symbol for a transformer in Fig. 5.11a does not give any indication of the phase of voltage across the secondary, since the phase of that voltage actually depends on the direction of the winding around the core. To solve this problem, the *dot notation* is used when the phase of the secondary is important. This notation is shown in Fig. 5.11b where the primary and secondary currents are in phase, and Fig. 5.11c where the currents are  $180^\circ$  out of phase.

## 5.8 VECTOR REPRESENTATION OF NO-LOAD CONDITIONS

Figure 5.12 is a vector diagram of conditions in a typical, unloaded step-down transformer. The flux lags the applied primary voltage by  $90^\circ$  and leads the induced primary and secondary voltages by  $90^\circ$ . The total no-load current is represented by  $I_E$ , and resolved into its core loss and magnetizing components by  $I_H$  and  $I_M$ , respectively. Since the core-loss current  $I_H$  is small in comparison with  $I_M$ , the magnetizing current is very nearly equal to the total primary current. Thus, the total no-load current is often called the *magnetizing current* or the *exciting current*. The cosine of the angle  $\theta$  by which the no-load current lags the applied voltage is the power factor of the transformer at no load. Power factor is given by

$$\text{Power factor} = \frac{I_H}{I_E} \quad (5.10a)$$

$$= \frac{I_H}{\sqrt{I_H^2 + I_M^2}} \quad (5.10b)$$

where  $I_H$  = core-loss current

$I_E$  = total no-load current

$I_M$  = magnetizing current

The total power expended in the transformer at no load is equal to  $EI \times \cos \theta$ . The

## 5-10 Transformers

copper, or  $I^2R$ , loss at no load is very small and need not be taken into account in most practical problems.

For the following problem draw a vector diagram approximating the conditions of the problem. Fill in the known values and then the unknown values as they are determined.

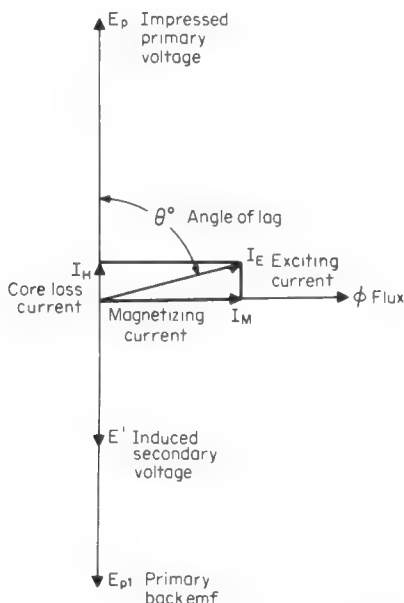


Fig. 5.12 Vector representation of no-load conditions.

**example 5.9** When the secondary of a 120- to 240-V transformer is open, the primary current from a 120-V source is 0.24 A at a power factor of 0.16. The transformer is rated at 2 kVA. Calculate: (a) the primary no-load exciting current in percentage of full-load current; (b) the core-loss current in amperes and in percentage of full-load current; (c) the magnetizing current in amperes and in percentage of full-load current.

**solution** (a) The transformer being rated at 2000 kVA, full-load primary current is

$$\frac{2000}{120} = 16.666 \text{ A}$$

$$\frac{\text{No-load primary current} \times 100}{\text{Full-load primary current}} = \% \text{ of full-load current}$$

Then, 
$$\frac{0.24 \times 100}{16.666} = 1.44\%$$

(b) Since the power factor is the cosine of the angle of lag between the exciting current and the core-loss current, core-loss current equals

$$\begin{aligned} I_E \cos \theta &= I_0 \times \text{PF} \\ &= 0.24 \times 0.16 = 0.0384 \text{ A} \end{aligned}$$

and

$$\frac{0.0384 \times 100}{16.666} = 0.23\%$$

(c) Reference to Fig. 5.12 will show that the magnetizing current  $I_m$  may be determined by use of the Pythagorean theorem. (The square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides.) In this case,

$$I_0^2 = I_E^2 + I_M^2$$

or

$$0.24^2 = I_M^2 + 0.0384^2$$

Solving for  $I_M$  results in

$$I_M = \sqrt{0.24^2 - 0.0384^2} = 0.2369 \text{ A}$$

and

$$\frac{0.2369 \times 100}{16.666} = 1.4214\% \text{ of full-load current}$$

In these results, it should be noted that the magnetizing current has nearly the same value as the no-load primary current. A vector diagram showing the conditions of this problem is given in Fig. 5.13.

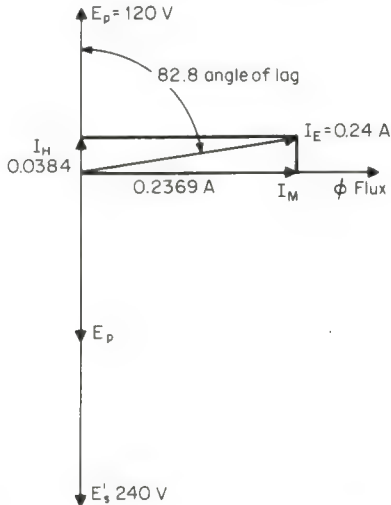


Fig. 5.13 Vector diagram of no-load conditions in the transformer of Example 5.9.

## 5.9 CONDITIONS UNDER LOAD

Figure 5.14 shows the transformer with a load. This illustration indicates the conditions existing at the moment when the terminal  $P_1$  of the transformer is positive, and is increasing. Application of the right-hand rule for coils shows that the magnetic field produced by the primary is then in the direction shown in the illustration. By Lenz's law, the current in the secondary must be in such a direction that its associated magnetic field must oppose the building up of a magnetic field by the primary current. In other words, the magnetic field of the secondary will oppose the field of the primary when the field due to the primary current is building up.

Also, by Lenz's law, the induced secondary voltage must be in a direction that opposes the change that produced it. Hence, when a load is connected to the secondary

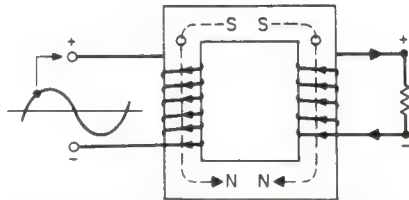


Fig. 5.14 As secondary load current increases, opposing magnetic field of secondary winding reduces inductance of primary, and more current flows into primary, thus maintaining core flux at its normal operating value.

## 5-12 Transformers

winding, the current flowing in the secondary produces a magnetomotive force which opposes, or tends to oppose, the flux  $\phi$  produced by the primary. Since an opposing flux, according to Eq. (5.6), results in a decrease in the back voltage of the primary, more current will flow into the primary as the secondary current increases! Therefore, any decrease of magnetization in the transformer core caused by current flowing in the secondary winding will immediately be counterbalanced by an increase in primary current. In this manner, the transformer automatically adjusts itself to load conditions, and the core flux is practically the same with full load as with no load.

Since the transformer is a highly efficient device, almost as much power is used up in the secondary as is fed to the primary. The power factor of the secondary, therefore, differs only slightly from that of the primary, and in most practical computations primary and secondary power factors may be considered equal to each other.

Because the core-loss and magnetizing current components of primary current are very much less than the load-current component, the core-loss and magnetizing current components may be added to the load-current component directly, instead of vectorially, except in those cases requiring the greatest accuracy. Such cases will be mostly of theoretical interest only.

**example 5.10** The no-load current taken by a 120- to 240-V transformer is 0.6 A. The transformer is rated at 2.4 kVA. Assuming that the primary and secondary power factors are equal to each other, determine the primary current when the secondary is supplying its rated kilovoltamperes to a load that has a power factor of 0.9.

**solution** Full-load secondary current:

$$I_s = \frac{2400}{240} = 10 \text{ A}$$

Since it is assumed that full-load primary and secondary power factors are equal, power factor may be disregarded in the calculations.

The transformer voltage ratio being 1 to 2, the primary component of the load current will be

$$2 \times 10 = 20 \text{ A}$$

Adding this directly to the 0.6-A no-load, or core-loss and magnetizing current component, the result is 20.6 A.

**example 5.11** Determine the ratio of full-load secondary power to full-load primary power for the transformer of Example 5.10.

**solution**

$$\begin{aligned}\text{Secondary power} &= 240 \times 10 \times 0.9 \\ &= 2160 \text{ W} \\ \text{Primary power} &= 120 \times 20.6 \times 0.9 \\ &= 2224.8 \text{ W}\end{aligned}$$

Ratio of full-load secondary power to full-load primary:

$$\frac{2400}{2242} = 0.9708$$

## 5.10 OPERATING A TRANSFORMER ON LOWER OR HIGHER THAN RATED FREQUENCY

Solving Eq. (5.8) for  $B_m$  (the maximum flux density in the core), we obtain

$$B_m = \frac{E_p \times 10^8}{4.44 \times f \times N_p \times A}$$

From this it may be seen that if we increase the voltage applied to the transformer primary, the flux will increase in proportion. The increased flux is, of course, due to the increased magnetizing ampere-turns, or magnetizing current.

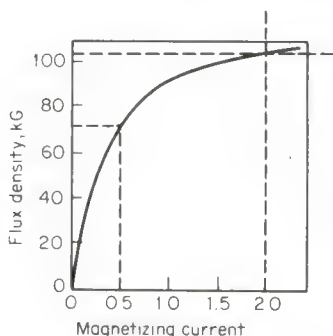
For efficient operation, it is necessary to operate a transformer with a *maximum* flux density below the knee of the core magnetization curve, as shown in Fig. 5.15. Below the knee of the curve, an increase in magnetizing current will be out of proportion to the resulting increase in  $B_m$ . For example, with a magnetizing current of 0.5 A, the flux density will be 70 kG. If the applied primary voltage is increased to raise the flux

density to 105 kG, the magnetizing current will rise to 2.0 A. Thus, as the flux density increased by 50 percent, the magnetizing current increased by 300 percent.

Equation (5.8) also shows that if the frequency  $f$  is decreased,  $B_m$  will again increase—and so will the primary current. Operating a transformer at considerably lower than rated frequency *could* cause the primary current to rise excessively, even to a value far in excess of the full-load primary current. The increased current would then burn out the winding.

From Eq. (5.8) it is evident that if the number of primary turns is increased,  $B_m$  decreases; but if the number of turns is decreased,  $B_m$  increases. Therefore, since  $B_m$  is directly proportional to the applied voltage and inversely proportional to the number of turns, changing the applied voltage and the number of turns in direct proportion would leave  $B_m$  and the primary current unchanged.

If we wanted to operate a transformer on a lower or higher than rated frequency, it would be impractical to change the number of its turns. However, since  $B_m$  is also directly proportional to the applied voltage and inversely proportional to the frequency,



**Fig. 5.15** Above the knee of the curve a large increase in magnetizing current results in a comparatively small increase in induction.

we could operate a transformer on lower than rated frequency, with normal flux, by decreasing the applied voltage in direct proportion to the decrease in frequency. For example, if we wished to operate a 120-V 60-Hz transformer at a frequency of only 25 Hz, it would be necessary to reduce the applied voltage to  $\frac{25}{60}$  of 120, or to 50 V. The transformer would then still operate at its normal flux and safe current levels, but its power output—and input—capacity and its core losses would all be lower in almost direct proportion to the decrease in operating frequency. It follows that if a transformer is to be operated at normal flux and current levels but at a higher than rated frequency, its primary applied voltage must be increased. Its power output capacity and core losses would, consequently, also increase.

From the above relations between operating frequencies and voltages, we get the proportion

$$F_1:F_2 = E_1:E_2 \quad (5.11)$$

where  $F_1$  = rated operating frequency of transformer

$F_2$  = changed operating frequency

$E_1$  = rated operating voltage of transformer

$E_2$  = changed operating voltage

A transformer may be safely operated at normal flux and current levels at any frequency if the ratio of applied voltage to frequency is kept at approximately its same, normal operating value.

**example 5.12** A 120-V 60-Hz transformer is to be operated at a frequency of 400 Hz. What voltage should be applied to the primary if the transformer is to operate at normal flux and current levels?

## 5-14 Transformers

**solution** By substitution in Eq. (5.11).

$$60:400 \approx 120:E_2$$

and solving for  $E_2$ ;

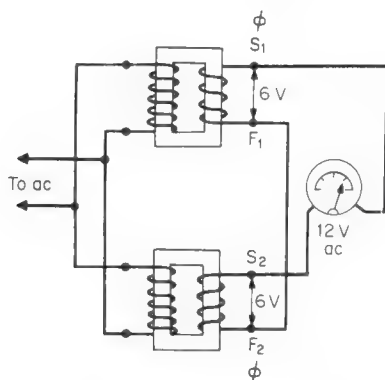
$$E_2 = \frac{120 \times 400}{60} = 800 \text{ V}$$

*Answer*

### 5.11 CONNECTING TRANSFORMERS IN SERIES AND PARALLEL

Occasions may often arise when it is necessary to connect two or more transformers in series or parallel, usually to obtain more current or more voltage than is available from only one transformer. It is essential that the transformers be properly phased when they are connected in series or parallel. Transformers may be damaged or ruined if they are connected together improperly.

**SERIES OPERATION** Connecting two transformers together to obtain a higher voltage is a very simple matter. It is only necessary to connect the primaries together in parallel, and connect their secondaries in series and to a voltmeter, as shown in Fig. 5.16. If the secondaries are properly phased, the meter will indicate the sum of



**Fig. 5.16** Phasing two transformers for series or parallel operation.

the secondary voltages. If the meter indicates the difference of the secondary voltages, the connection to one of the secondaries may be reversed, or one of the primary windings may be reversed.

**PARALLEL OPERATION** If two transformers are to be connected in parallel, the transformers should be alike in every respect. They should have the same number of primary and secondary turns and the same secondary voltages, and the secondary voltages should remain the same under load. If the secondaries of two transformers having unequal voltages were to be connected in parallel, the transformer with the higher secondary voltage would, in addition to supplying its load, also send a current through the secondary winding of the other transformer. The result would be a large copper ( $I^2R$ ) loss in the windings, which might produce enough heat to damage the winding insulation.

Even though two transformers may have the same primary and secondary voltages, they may differ in other respects. One winding may have a greater internal drop than the other when the transformers are supplying a load, and this would cause a difference in their terminal voltages under load.

The usual reason for connecting two transformers for parallel operation is that one does not have the necessary current-delivering capacity.

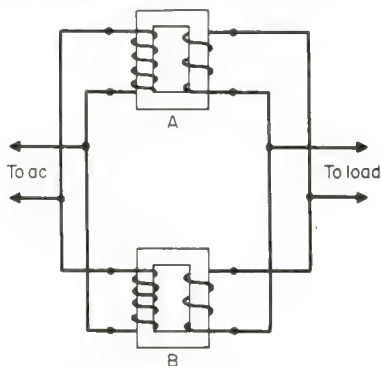
**PHASING TRANSFORMERS FOR PARALLEL OPERATION** If the secondaries of two transformers are not properly phased for parallel operation, the result is a short



circuit that burns out the windings. Even if two transformers of the same make have the same numbered or lettered terminals, it should not be assumed that they will be properly phased for parallel operation.

To phase two transformers for parallel operation, we connect the primaries in parallel and the secondaries in series—just as we would in phasing the transformers for series operation—as shown in Fig. 5.16. The secondary terminals of one of the transformers are arbitrarily tagged  $S_1$  and  $F_1$  and those of the other transformer  $S_2$  and  $F_2$ . If the voltmeter across the terminals,  $S_1$  and  $S_2$ , reads twice the voltage of either transformer, we connect terminal  $S_1$  to  $F_2$  and terminal  $F_1$  to  $S_2$ . However, if the voltmeter reading is zero,  $S_1$  and  $F_2$  have opposite polarities, and we connect  $S_1$  to  $S_2$  and  $F_1$  to  $F_2$ .

The phasing of transformers with high-voltage secondaries or primaries should be done with lowered excitation voltages in order to reduce high-voltage shock hazard.



**Fig. 5.17** What will be the result if connections to one of the primaries are reversed?

**example 5.13** Two transformers are connected together for parallel operation, as in Fig. 5.17. If the connections to the primary of transformer A are accidentally reversed, what will be the result?

**solution** The polarity of the secondary winding of transformer A will also be reversed, and the result will be a short circuit of the secondary windings.

## 5.12 TRANSFORMER LOSSES

Power losses in a transformer occur in the transformer core material and in the transformer primary and secondary windings.

Eddy-current losses, as explained in Chap. 4, are the result of currents circulating in the core and producing heat. The circulating current in the core is set up by induced voltages in the core material. These voltages are the result of the changing magnetic flux.

Hysteresis loss is the energy lost by reversing the magnetic field in the core as the magnetizing alternating current rises and falls and reverses direction. Hysteresis loss was also described in Chap. 4.

**COPPER LOSS** Copper loss is the power lost in the primary and secondary windings of the transformer. It is due to the ohmic resistance of the windings. Copper loss is obtained by the equation

$$\text{Copper loss} = I_p^2 R_p + I_s^2 R_s$$

where  $R_p$  = resistance of primary

$R_s$  = resistance of secondary

$I_p, I_s$  = respective primary and secondary currents

**DETERMINATION OF COPPER LOSS** Copper loss may be determined by measuring the ohmic resistance of each winding and then computing the  $I^2 R$  loss in both primary and secondary windings for any given secondary load current.

## 5-16 Transformers

The copper loss in both windings may also be determined by means of a wattmeter, and a circuit diagram showing the proper connections for this method is given in Fig. 5.18. If a transformer has multiple primaries, secondaries, or both, all the coils must be connected into the circuit for this test.

At the start of the test, the variable-voltage autotransformer is set at zero; its setting is then gradually raised until the ammeter connected across the secondary indicates rated full-load secondary current (assuming that we wish to determine copper loss at full load). At this point the wattmeter will indicate the power loss in both windings and also a very small core loss. Usually the core loss is too low to be indicated on the wattmeter.

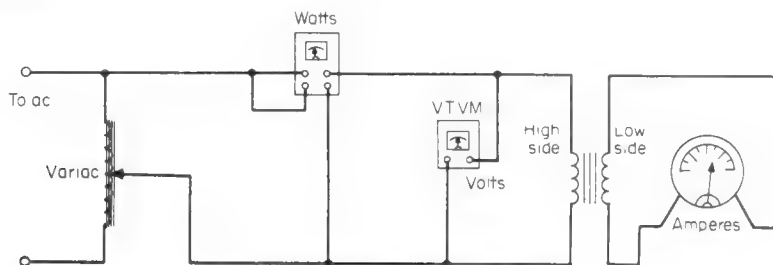


Fig. 5.18 Short-circuit test for copper loss.

Although rated primary and secondary currents flow through the windings when this method is used, core loss is negligible because the voltage impressed on the primary has only a very low value—perhaps less than 5 percent of normal. This low primary voltage is still enough to cause rated secondary current to flow in the secondary winding, which is practically short-circuited by the ammeter connected across its terminals. Since the transformer is operating at only about five percent of rated voltage, the maximum flux density  $B_m$  in the core is also only about five percent of normal. Since core loss is almost directly proportional to  $B_m^2$ , the core loss with five percent of normal impressed primary voltage will be  $0.05 \times 0.05$  or  $0.0025$  ( $25_{100}$  of 1 percent) of normal. As an example, if normal core loss in a transformer is 50 W, the loss on short circuit with five percent of rated primary voltage will be  $0.05^2 \times 50$ , or 0.125 W, a value too small to be noticeable on the wattmeter.

The above test for copper loss is known as the *short-circuit test* method. Copper loss by this method may be determined from either the primary or secondary side of the transformer; the high side will usually be more convenient because on short circuit it would take only about five percent of full-rated voltage to cause full-load current to flow through the windings, but it might not always be convenient to adjust the applied voltage to the proper low value needed for a test from the low side.

**CORE-LOSS TEST** Core loss in a transformer may be determined by applying the rated voltage to the transformer primary at its rated frequency and measuring the power input with the secondary open-circuited. Power input is measured by means of a wattmeter. Figure 5.19 shows a diagram of the necessary connections. The variable voltage autotransformer is adjusted to apply rated voltage to the transformer primary. This will be indicated by the VTVM (or other high-resistance meter) connected across the primary. (If a low-resistance voltmeter were to be used to measure the applied voltage, its power consumption would be added to the core-loss power, and the wattmeter would then give a false core-loss reading. Therefore, if a low-resistance voltmeter is used to measure the applied voltage, the wattmeter reading should be taken with the circuit to the voltmeter open. The input power measured by the wattmeter will include a small copper loss, but at no load this is of no consequence and may be ignored.

If the rated primary voltage of the transformer has a very high value, or whenever it is higher than the secondary voltage, core loss may be determined by applying rated secondary voltage to the transformer. This is called the *open-circuit test*, and it is

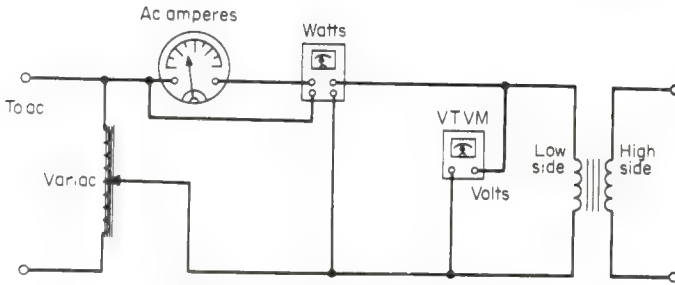


Fig. 5.19 Open-circuit test for core loss.

usually more convenient and safer for any step-down transformer. It gives the same results as a core-loss measurement taken on the high or primary side of the transformer.

**example 5.14** A 240- to 24-V step-down transformer has a full-load secondary current rating of 50 A. A short-circuit test for copper loss at full load gives a wattmeter reading of 100 W. If the resistance of the primary is  $0.8\ \Omega$ , what is the resistance of the secondary, and what is the secondary power loss? (Ignore the small core loss.)

**solution** Since full-load secondary current is 50 A, and the voltage ratio of the transformer is 240 to 24, or 10 to 1, the full-load primary current is  $\frac{50}{10}$  or 5 A. Total copper loss equals

$$100\text{ W} = I_p^2 R_p + I_s^2 R_s = 5^2$$

Then,

$$\begin{aligned} R_s &= \frac{100 - 5^2(0.8)}{50^2} = \frac{100 - 20}{2500} \\ &= \frac{80}{2500} = 0.032\ \Omega \end{aligned}$$

Secondary power loss is

$$50^2 \times 0.032 = 80\text{ W}$$

**example 5.15** On an open-circuit test for core loss, the transformer of Example 5.14 takes 1.2 A from a 240-V ac source. If the wattmeter reading is 80 W, determine (a) the copper loss at zero load; (b) the core loss.

**solution** (a) Since the primary resistance is  $0.8\ \Omega$ , and the secondary is an open circuit, copper loss occurs only in the primary and is equal to

$$I_p^2 R_p = 1.2^2 \times 0.8 = 1.15\text{ W}$$

(b) The wattmeter reading minus the copper loss equals the core loss, or

$$80 - 1.15 = 78.8\text{ W}$$

### 5.13 EFFICIENCY

The efficiency of a transformer, like the efficiency of any machine, is the ratio of power output to power input, or in the form of an equation:

$$\begin{aligned} \text{Efficiency} &= \frac{\text{power output}}{\text{power input}} \\ &= \frac{\text{power output}}{\text{power output} + \text{copper loss} + \text{core loss}} \\ &= \frac{E_s I_s \times \text{PF}}{(E_s I_s \times \text{PF}) + \text{copper loss} + \text{core loss}} \end{aligned} \quad (5.12)$$

where PF = power factor of the load.

It might be assumed that transformer efficiency may be determined by connecting wattmeters in the primary and secondary of a transformer—that is, supplying a load, and then dividing the secondary wattmeter reading by that of the primary. However, this would be an incorrect assumption. Transformer efficiency is very high, and there-

## 5-18 Transformers

fore the copper and core losses are very low. A wattmeter connected in the secondary circuit of a transformer would read very nearly the same as a wattmeter connected in the primary, and because of meter inaccuracies and errors introduced in reading the meters, it would be impossible to determine power input and output with the degree of precision necessary for computing transformer efficiency.

To compute efficiency, core loss must be determined. This can be done by the core-loss test method described earlier in this chapter. Copper loss may be determined by the measured resistance method, or by the short-circuit test. These core and copper loss factors are substituted in Eq. (5.12), which will then give the efficiency.

**example 5.16** An open-circuit test for core loss in a 240- to 720-V 5-kVA transformer gives a value of 60 W. The measured resistance of the low side winding (primary) is  $0.025 \Omega$ , and that of the high side (secondary) is  $1.25 \Omega$ . If the power factor of the load is 0.84, what is the efficiency of the transformer at full load?

**solution** Full-load secondary current  $5000/720$ ; and full-load secondary copper loss,

$$\left(\frac{5000}{720}\right)^2 1.25 = 60.28 \text{ W}$$

Full-load primary current  $5000/240$ ; and full-load primary copper loss,

$$\left(\frac{5000}{240}\right)^2 \times 0.025 = 10.85 \text{ W}$$

Total copper loss,

$$60.28 + 10.85 = 71.1 \text{ W}$$

Efficiency,

$$\frac{5000 \times 0.84}{(5000 \times 0.84) + 60 + 71.1} = 0.971 \text{ or } 97.1\%$$

## 5.14 EQUIVALENT RESISTANCES

In certain computations it is sometimes useful to assume that the copper loss in both primary and secondary windings of a transformer actually occurs in only the primary; this is equivalent to assuming that the wattmeter in a short-circuit test for copper loss actually indicates power being expended in a purely resistive load. In the form of an equation this may be written as

$$W = I_p^2 R_{ep}$$

or

$$R_{ep} = \frac{W}{I_p^2} \quad (5.13)$$

where  $W$  = wattmeter reading on a short-circuit test

$I_p^2$  = primary current squared

$R_{ep}$  = equivalent resistance of transformer

It is important to note that  $R_{ep}$  in this case is *not* the dc resistance of the primary, but merely an assumed resistance that would give the same  $IR$  drop as the *combined*  $IR$  drops of both primary and secondary windings.

It is also possible to express the equivalent primary resistance in terms of the resistances of both primary and secondary windings in the form

$$R_{ep} = R_p + R_s \left(\frac{N_p}{N_s}\right)^2 \quad (5.14)$$

And it is possible to express the *equivalent secondary resistance*  $R_{es}$  also in terms of both primary and secondary resistances, as follows:

$$R_{es} = R_s + R_p \left(\frac{N_s}{N_p}\right)^2 \quad (5.15)$$

From Eqs. (5.14) and (5.15) we also get

$$\frac{R_{ep}}{R_{es}} = \frac{N_p^2}{N_s^2} \quad (5.16)$$

**example 5.17** A 1200- to 120-V transformer is rated at 5 kVA. The primary resistance is 1  $\Omega$ , and secondary resistance is 0.05  $\Omega$ . Determine (a) the equivalent primary resistance  $R_{ep}$ ; (b) the equivalent secondary resistance  $R_{es}$ ; (c) the total copper loss  $I^2R$ , using the determined equivalent secondary resistance; (e) the total copper loss, using the dc resistances of the primary and secondary windings.

**solution** (a) The equivalent primary resistance,

$$R_{ep} = 1 + 0.05 \left( \frac{N_p}{N_s} \right)^2$$

Since the voltage ratio of the transformer is 1200/120, or  $10/1$ ,  $N_p/N_s$  also equals  $10/1$ , or 10. Therefore

$$R_{ep} = 1 + 0.05(10)^2 \\ = 6 \Omega$$

(b) The equivalent secondary resistance,

$$R_{es} = 0.05 + 1 \left( \frac{N_s}{N_p} \right)^2 \\ = 0.05 + (1/10)^2 = 0.06 \Omega$$

(c) Primary current,

$$I_p = \frac{5000}{1200} = 4.175 \text{ A}$$

Copper loss computed from  $R_{ep}$  and  $I_p$  values equals

$$4.175^2 \times 6 = 104.58 \text{ W}$$

(d) Secondary current,

$$I_s = \frac{5000}{120} = 41.75 \text{ A}$$

Copper loss computed from  $R_{es}$  and  $I_s$  values equals

$$41.75^2 \times 0.06 = 104.58 \text{ W}$$

(e) Primary copper loss equals

$$(I_p^2 R_p) = 4.175^2 \times 1 = 17.43 \text{ W}$$

Secondary copper loss equals

$$(I_s^2 R_s) = 41.75^2 \times 0.05 = 87.15 \text{ W}$$

Total copper loss,

$$I_p^2 R_p + I_s^2 R_s = 17.43 + 87.15 = 104.58 \text{ W}$$

## 5.15 VOLTAGE REGULATION

The voltage regulation of a transformer in percent of the rated full-load secondary voltage is given by the equation

$$\text{Percent regulation} = \frac{\text{no-load secondary volts} - \text{full-load secondary volts}}{\text{full-load secondary volts}} \times 100 \quad (5.17)$$

The voltage regulation of small transformers may easily be determined by adjusting the primary voltage so that the transformer delivers its rated kilovoltamperes at its rated secondary voltage to a load at some specified power factor. Then, disconnecting the load but keeping the primary voltage constant will cause the secondary voltage to rise to its no-load value. Substituting the observed values of full-load and no-load voltages in Eq. (5.17) will give the percent regulation.

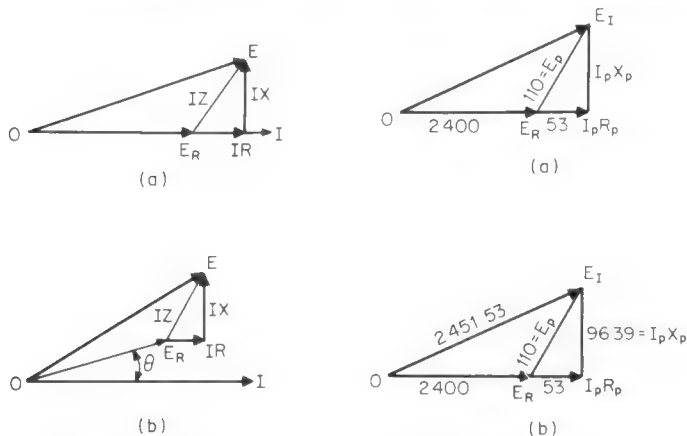
The above method, however, would not be very convenient for large heavy-duty transformers, especially when such transformers may also have high-voltage primary or secondary windings. For such transformers it is better to compute regulation from the short-circuit test data. The computation for a load having either a lagging or leading power factor is somewhat more complicated than for a load having unity power factor.

Regulation may be computed in terms of either the primary or secondary winding,

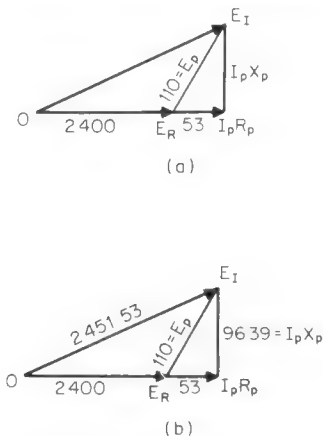
## 5-20 Transformers

and the first step, in either case, is to compute the equivalent  $IR$  and  $IX$  drops for the winding from the short-circuit test data. Determining regulation in terms of the secondary winding is equivalent to assuming that the secondary is really the primary; and regardless of whether we compute regulation in terms of either the primary or secondary winding, the no-load voltage will always be greater than the rated full-load voltage of the transformer.

Vector diagrams for computing regulation from known values of  $E_R$ ,  $I_p$ , and  $IX$  are shown in Fig. 5.20. A vector diagram for a transformer supplying a unity power factor load is shown in Fig. 5.20a and for a load with a lagging power-factor in Fig. 5.20b. The vector  $OE_R$  represents the rated full-load voltage (primary or secondary) of the transformer. Vectors  $IR$ ,  $IX$ , and  $IZ$  are the equivalent (primary or secondary) resistance, reactance, and impedance drops. Vector  $OE_I$  is the vector sum of  $E_R$  and  $IZ$ , and it represents the no-load voltage of the transformer. In (a) the current  $I$  is in phase with  $E_R$ , since the load has a power factor of unity. In (b) the current  $I$  lags  $E_R$  by the angle  $\theta$ , corresponding to the power factor (cosine of the angle of lag) of the load. In



**Fig. 5.20** Vector diagrams for computing regulation from known values of  $E_R$ ,  $I_p$ , and  $IX$ .  $E_R$  equals the rated voltage of the transformer.  $IR$ ,  $IX$ , and  $IZ$  are the equivalent values of resistance, reactance, and impedance. (a) Unity power-factor load. (b) Lagging power-factor load.



**Fig. 5.21** Vector diagrams for Example 5.18. Computing  $OE_I$ , the no-load voltage, is a problem in simple trigonometry. (a) Partially solved problem. (b) Completed problem.

either case, the problem is to find the value of the vector  $OE_I$ , the no-load voltage of the transformer, from the known values of  $E_R$ ,  $I_p$ , and  $IX$ .

**example 5.18** On a short-circuit test, a 2400- to 240-V, 10 kVA transformer takes 4.17 A at an applied primary voltage of 110 V. The wattmeter reading is 221 W. Assuming a unity power factor load, compute the regulation in terms of the primary. Draw a vector diagram illustrating the conditions of the problem, and fill in the unknown values as they are calculated.

**solution** Since the power factor of the load is unity, the applied voltage and the current will be in phase. A vector diagram of circuit conditions will, therefore, be similar to the one shown in Fig. 5.20a. The voltmeter reading is the equivalent primary impedance times the primary current:

$$I_p Z_{ep} = 110 \text{ V}$$

The equivalent primary resistance in volts is

$$I_p R_{ep} = \frac{W}{I_p} = \frac{221}{4.17} = 53 \text{ V}$$

Figure 5.21 shows the vector diagrams for this problem. Figure 5.21a illustrates the conditions of the problem at this stage. By simple trigonometry,

$$(I_p X_p)^2 = 110^2 - 53^2$$

Hence the equivalent primary reactance in volts is

$$I_p X_p = \sqrt{110^2 - 53^2} = 96.39 \text{ V}$$

Since  $E_R$  is in phase with  $I_p R_p$ , we may add  $E_R$  to  $I_p R_p$  directly, or

$$E_R + I_p R_p = 2400 + 53 = 2453$$

and again using simple trigonometry,

$$\begin{aligned} (OE_I)^2 &= 2453^2 + (I_p X_p)^2 \\ &= 2453^2 + 96.39^2 \end{aligned}$$

Solving for the no-load voltage

$$\begin{aligned} OE_I &= \sqrt{2453^2 + 96.39^2} \\ &= 2451.53 \text{ V} \\ \text{Regulation} &= \frac{2451.53 - 2400}{2400} \times 100 = 2.147\% \end{aligned}$$

Figure 5.21*b* is a vector diagram of the completed problem.

## 5.16 THE ISOLATION TRANSFORMER

An isolation transformer is used whenever it is necessary to avoid a direct electric connection between a piece of electric equipment and the power lines or other source of power. This is useful and very often even necessary where equipment must be grounded for increased efficiency, to reduce noise, or for other reasons. Since one side of the ac power line is grounded, connecting a piece of grounded equipment to the lines could result in a short circuit unless precautions are taken to assure that the grounded side of the equipment is connected to the grounded side of the power lines. There are also certain types of equipment—such as those with ac-dc power supplies—in which a lethal shock *could* be received when the chassis is exposed. (It would be necessary to touch a ground point *and* the exposed chassis to receive such a shock.) By using an isolation transformer, the chassis becomes a floating ground, and it is nearly impossible to receive a deadly shock under such conditions.

Since equipment may frequently be used in different locations and because power outlets in the same building or even in the same room are not usually polarized, a transformer between the equipment and the power lines is a very convenient and practical method of avoiding the possibility of damaging a piece of electric equipment that must be grounded. Since the transformer serves to isolate the equipment from the power lines, it is called an isolation transformer; and it is evident that any transformer may be used as an isolation transformer if it has the necessary voltampere rating and the required primary to secondary turns or voltage ratio.

## 5.17 THE AUTOTRANSFORMER

In an ordinary transformer the primary and secondary windings are magnetically coupled, but there is no direct electric connection between the primary and secondary windings. In an autotransformer, the primary and secondary windings are not only magnetically coupled, but there is also a direct electric connection between the windings. In fact, the secondary winding may be merely a continuation of the primary.

Any ordinary transformer may be used as an autotransformer. For example, consider the 1-to-1 isolation transformer of Fig. 5.22 in which there is a direct series connection between the primary and secondary windings. With the transformer phased as shown, the impressed primary voltage is added to the induced secondary voltage, and the total voltage across the load, connected to terminals *A* and *Y*, is

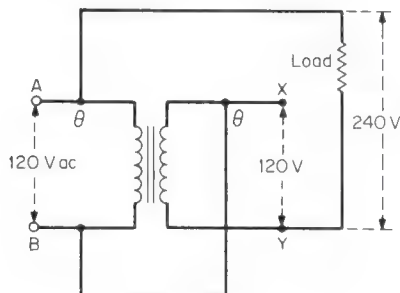
$$120 + 120 = 240 \text{ V}$$

Thus we have a step-up ratio of 1 to 2 in a transformer with a primary to secondary turns ratio of only 1 to 1. Figure 5.23 shows some additional connections of autotransformers in which a transformer with a 1-to-1 ratio is used to step up the voltage (Fig. 5.23*a*), and to step down the voltage (Fig. 5.23*b*). In Fig. 5.23*a* the windings of the transformer are wound on a straight, open core, and this more clearly illustrates the conditions of the transformer of Fig. 5.22. This same transformer may also be used as a step-down auto-

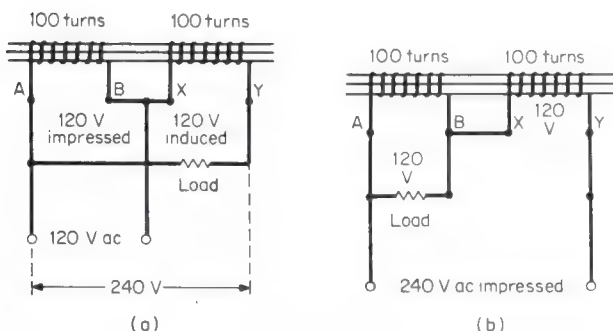


## 5-22 Transformers

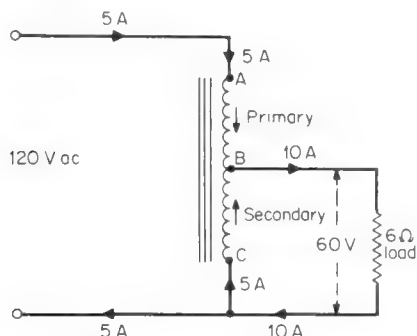
transformer by connecting the load across terminals *A* and *B* and connecting terminals *A* and *Y* to a power source, as shown in Fig. 5.23*b* which merely interchanges the line and load connections of Fig. 5.23*a*. In other words, the transformer of Fig. 5.23*a* is operated in reverse—just as we might operate any transformer in reverse by using the secondary, with rated applied voltage, as the primary and then connecting the load to the normal primary.



**Fig. 5.22** A 1-to-1 isolation transformer, connected for use as an autotransformer, gives a stepped-up voltage ratio of 1 to 2.



**Fig. 5.23** Some additional connections of autotransformers: (a) Auto-transformer connection steps up impressed primary voltage from 120 to 240 V. (b) Same 1:1 ratio transformer used in reverse steps down impressed line voltage of 240 to 120 V.



**Fig. 5.24** Autotransformer supplying a load. Current transformed is 5 A, and in this case the current transferred is also 5 A.

A schematic diagram of an autotransformer supplying a load is shown in Fig. 5.24. In this case we have a 2-to-1 step-down transformer. The number of turns between A and B is the primary, and the turns between B and C the secondary. If we assume that the transformer is 100 percent efficient and supplies a load having unity power factor, power input to the transformer will be

$$120 \times 5 = 600 \text{ W}$$

Power output will be

$$60 \times 10$$

which also equals 600 W, but the transformer primary power is only  $60 \times 5$ , or 300 W. This is also the transformer secondary power. Hence, the power *transformed*, by transformer action, is only 300 W; and the power *transferred* directly to the secondary load, by means of the connection between the primary and secondary windings, is the remaining 300 W, making the total of 600 W supplied to the load.

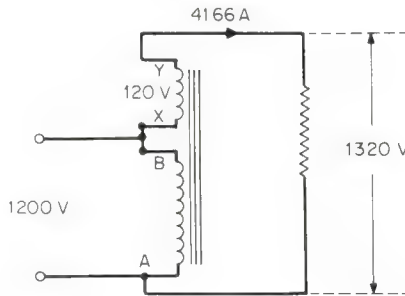
Since, in this case, each half of the autotransformer supplies five amperes to the load, copper loss is only one-half the copper loss that would occur in a transformer with separate primary and secondary windings and supplying the same load with the same current.

As the primary to secondary voltage ratio of an autotransformer decreases, the copper loss decreases in proportion, and the amount of power *transferred* directly—instead of by transformer action—increases in proportion.

Autotransformers are seldom used when it is necessary to reduce a very high voltage to a much lower value. For example, a voltage ratio of 1200 to 120, or 10 to 1, would result in a small reduction of copper loss, and the power *transferred* directly would also be small. Furthermore, a dangerous shock hazard would exist in the secondary, since there would be no isolation between the high-voltage primary and the low-voltage secondary.

If an ordinary transformer having several windings is to be used as an autotransformer by connecting all the windings together, the current carried by any winding must not be greater than its normal rated current.

**example 5.19** The windings of an ordinary 1200- to 120-V 5-kVA transformer are connected in series to form a step-up autotransformer, as shown in Fig. 5.25. Assuming that the



**Fig. 5.25** Ordinary transformer connected as autotransformer; in this case power capacity is increased for 5000 to 55 000 W.

transformer is 100 percent efficient and operates at a power factor of unity with rated secondary current, determine (a) the power delivered to the load; (b) the total current supplied by the source; (c) the increase in the power rating of the transformer when it is used as an autotransformer; (d) the power transferred.

**solution** (a) Since rated secondary current is

$$\frac{5000}{120} = 41.66 \text{ A}$$

## 5-24 Transformers

the current in the load may also be 41.66 without overloading the secondary winding, XY. Also, since the voltage across the load is  $1200 + 120$  or 1320 V, the power supplied to the load is

$$1320 \times 41.66 = 55\,000\text{ W}$$

(b) Since we have assumed 100 percent efficiency, power supplied by the source is equal to the power delivered; current taken from the source is, therefore,

$$\frac{55\,000}{1200} = 45.83\text{ A}$$

(c) Normal power rating of the transformer is 5000 W. Power rating of the transformer when used as an autotransformer is 55 000 W. Percentage increase in power rating,

$$\frac{55\,000}{5000} \times 100 = 1100\%$$

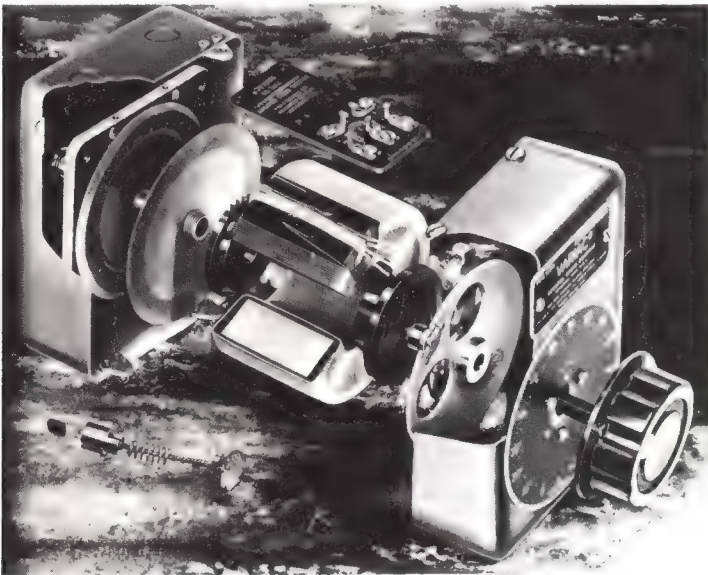
(d) Power transferred is

$$55\,000 - 5000 = 50\,000\text{ W}$$

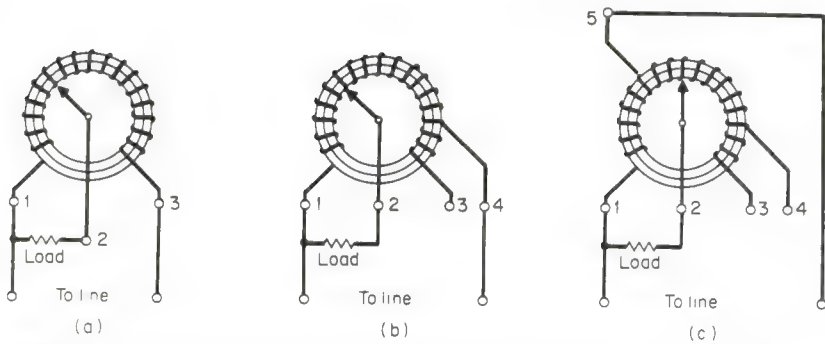
## 5.18 THE VARIABLE AUTOTRANSFORMER

The variable autotransformer is usually better known under one of its commercial trade names, such as Variac, Powerstat, Adjust-A-Volt. But regardless of what we call it, its basic principle of operation remains the same. Figure 5.26 is a photograph of a commonly used Variac. The exploded view clearly shows all the basic parts. Figure 5.27 shows several autotransformer circuits.

Figure 5.27a illustrates the most basic type of variable autotransformer. The coil is a single-layer winding over a toroidal core of high permeability. The sliding brush contact, which is controlled by its attached knob or dial, may be set to tap the winding at any point about its periphery. Thus, the output voltage across the load, connected to terminals 1 and 2, may be varied from zero to full line voltage. A slight modification of the design, as shown in Fig. 5.27b, permits the voltage to be varied from zero to slightly (approximately 17 percent) above line voltage. This feature is useful when-



**Fig. 5.26** An exploded view of a Variac. The brush, shown at lower left, would normally contact the front outer perimeter of the winding. (Courtesy General Radio Company.)



**Fig. 5.27** Connections for autotransformers: (a) The basic variable autotransformer. (b) Tapped winding permits adjustment from zero to slightly above line voltage. (c) Tapping the winding at a lower point permits adjustment from zero to more than twice the line voltage.

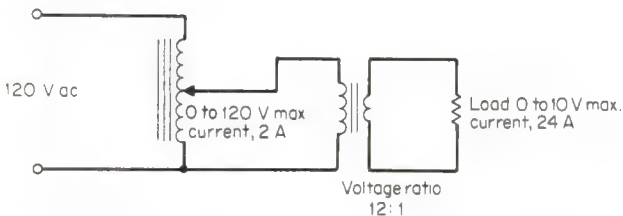
ever it is necessary to compensate for the drop in line voltage that occurs when the power lines are heavily loaded. But if necessary, this design may also be connected for a range of zero to line voltage by joining terminals 1 and 3, instead of 1 and 4, to the lines.

An additional modification of the basic design, as shown in Fig. 5.27c, gives a voltage range adjustable from zero to slightly more than twice the line voltage (from 0 to 280 V for one model). Terminal 5 is connected to a tap on the winding at a point slightly less than midway between the total number of turns.

The rated current of a variable autotransformer should not be exceeded at any voltage setting of the transformer, and it should not be assumed that greater than rated current may safely be drawn at voltage settings below line voltage—as we might use an ordinary step-down transformer to supply a much greater current to a secondary load than the current supplied to the primary.

When the variable autotransformer is used to supply a fixed, known load, without using the overvoltage connection, maximum current may be supplied to the load at full line voltage. Then as the voltage setting is reduced, the decrease in current with decreasing output voltage will tend to keep the load current within safe limits for any voltage setting. Accidental overload damage may easily be avoided by fusing the variable autotransformer for rated current.

When it is necessary to supply a load at some reduced voltage but with a current considerably in excess of the normal current rating of a particular variable autotransformer, an auxiliary transformer, connected to the output terminals of the variable autotransformer, may be used, as shown in Fig. 5.28. For the transformer and circuit of Fig. 5.28, the primary-to-secondary voltage ratio of the transformer is 12 to 1, and normal rated current of the variable autotransformer is two amperes. Current taken by the load may, therefore, be as great as 24 A at voltages between 0 and 10 V. In this case, the use of an auxiliary transformer has increased current capacity 12 times. In any application of this

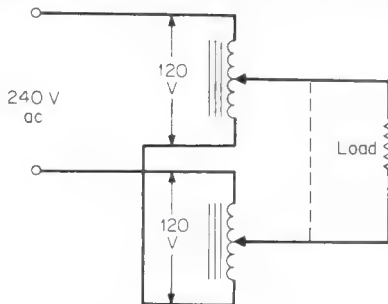


**Fig. 5.28** The use of supplementary transformer in conjunction with low-current variable autotransformer increases current capacity 12 times.

sort, the voltampere rating of the transformer should be equal to the voltampere rating of the variable autotransformer.

### USING VARIABLE AUTOTRANSFORMERS AT HIGHER THAN RATED VOLTAGES

A single variable autotransformer should not, of course, be operated at higher than normal rated voltage. However, two variable autotransformers may be ganged for high-voltage operation, as shown in Fig. 5.29. Thus, if the normal voltage rating of each



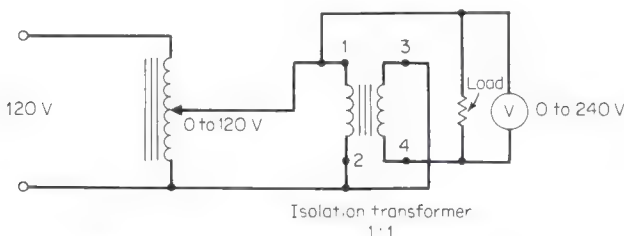
**Fig. 5.29** Two 120-volt variable autotransformers ganged for operation on 240 V.

transformer is 120 V, the ganged transformers may be used at 240 V, or two 240-V models may be used at 480 V.

**SWITCHING** Before a variable autotransformer is switched into or out of a circuit, its setting should be reduced to zero to prevent excessive surge currents from damaging the windings. This is especially important when a variable autotransformer is used to supply a load that has a very high ratio of hot-to-cold resistance. For example, the hot-to-cold resistance ratio of the heaters or filaments of vacuum tubes or of incandescent lamps may be as great as 15 to 1.

**example 5.20** A 120-V 1-to-1 isolation transformer is available. Draw a diagram showing how this transformer should be connected to a 120-V Variac and to a load that is to be operated at various voltages ranging from 0 to 240 V.

**solution** The primary winding of the transformer is connected to the Variac, and then the transformer itself is connected to form an autotransformer, as shown in Fig. 5.30. This gives a



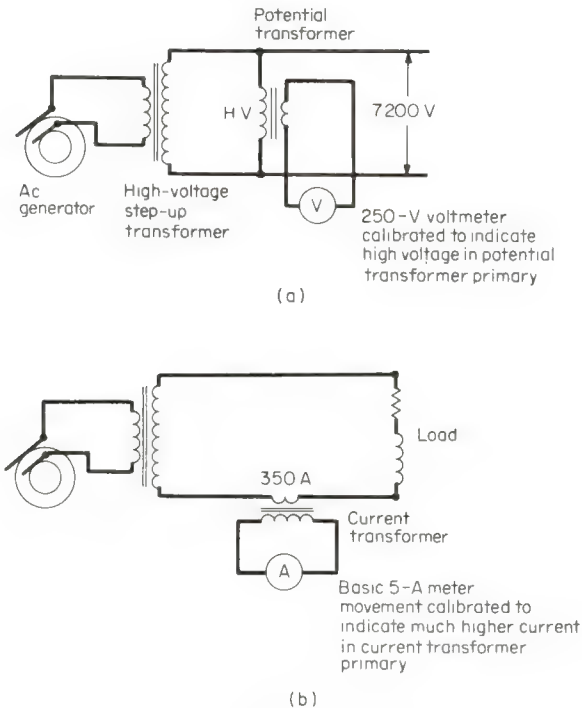
**Fig. 5.30** Increasing the voltage range of a Variac.

step-up ratio of 1 to 2. If the voltmeter indicates zero as the Variac setting is increased, the connections to either terminals 1 and 2 or 3 and 4 should be reversed — for proper phasing.

### 5.19 INSTRUMENT TRANSFORMERS

Instrument transformers are designed for use with measuring instruments, relays, or control devices. They are especially designed to maintain a specific relationship in magnitude and phase between the primary and secondary currents or voltages.

**THE POTENTIAL TRANSFORMER** Since it would hardly be feasible to construct voltmeters to directly measure the extremely high voltages encountered in power distribution systems, such high voltages cannot be measured by connecting a voltmeter directly across the lines. Even if voltmeters having the necessary range and insulation resistance were available, it would be very dangerous to bring leads from the high-voltage lines directly into the instrument switchboards and to the meters. A dangerous high-voltage shock hazard would then be a constant threat to personnel at the switchboard. Therefore, instead of bringing the high-voltage leads directly to a meter, these leads are connected to one side of a small transformer, called a potential transformer, as shown in Fig. 5.31. In Fig. 5.31a the potential transformer steps down the high volt-



**Fig. 5.31** Instrument transformers connected to measure high voltage and current: (a) Potential transformer. (b) Current transformer.

age to a very much lower value that may be safely brought to a low-voltage meter. Usually, the meter has its scale calibrated to indicate the high voltage across the transformer's primary terminals. Since the potential transformer has very high primary-to-secondary insulation resistance, there is no high-voltage shock hazard at the meter terminals.

**THE CURRENT TRANSFORMER** A current transformer, instead of being used to step down a high voltage, is used to step down a high current to a much smaller and more conveniently measurable value.

The method of connecting a current transformer into the line is shown in Fig. 5.31b. Since a current transformer is always connected in series with the line, it is also known as a *series* transformer. The primary, or line side, of a current transformer usually has only a few turns—sometimes only one—since it must carry a very large current. The secondary, which is connected to a low-range ammeter, has a much larger number of turns. The meter scale is calibrated to indicate the line current in the primary of the

## 5-28 Transformers

current transformer. This line current is, of course, very much greater than the current in the secondary of the current transformer.

Although the use of a current transformer completely insulates the secondary and the ammeter from the high-voltage lines, the secondary of a current transformer should never be left open-circuited. To do so may result in a dangerous high voltage being induced in the secondary. Therefore, if on any occasion it should be necessary to disconnect the meter used with a current transformer, the transformer secondary terminals should first be short-circuited, and this can do no harm because the primary current will still be limited to the amount that the load will permit to flow.

**example 5.21** The voltage across the secondary of a potential transformer is 180 V. The primary, or line side, of the transformer has 24 000 turns, and the secondary has 600 turns. What is the line voltage?

**solution** Dividing 24 000 by 600 shows that the turns or voltage ratio of the transformer is 40 to 1. Then, since the secondary voltage is 180, the primary or line voltage is

$$40 \times 180 = 7200 \text{ V}$$

**example 5.22** The current ratio of a current transformer is 100 to 1. What is the line current when an ammeter connected to the secondary indicates 3.5 A?

**solution** The current ratio being 100 to 1, the line current is

$$100 \times 3.5 = 350 \text{ A}$$



# Chapter 6

## Practical Circuit Analysis

### 6.1 INTRODUCTION

Electronic technicians are often required to analyze simple circuits energized by dc or ac sources. Generally, only elementary algebra and trigonometry are required in their analysis. The object of this chapter is to review some basic circuit laws and their application in the analysis of dc and ac circuits. Topics discussed are illustrated by pertinent examples and their complete solution.

### 6.2 NOTATION

By convention, for time-varying quantities, and often in general discussions, *lowercase* letters are used. Voltage is designated by  $e$ , current by  $i$ , energy by  $w$ , and power by  $p$ . For specific dc and ac quantities, *uppercase* letters are used: voltage  $E$ , current  $I$ , energy  $W$ , and power  $P$ .

### 6.3 DEFINITIONS

We first define some basic terms used in the analysis of circuits.

**PASSIVE AND ACTIVE ELEMENTS** An element is considered *passive* if it is capable of only *absorbing* electric energy. A resistor, an inductor, and a capacitor are passive elements. Elements that *provide* electric energy are *active elements*. Examples of active elements include the battery (used to energize circuits) and the transistor (which may be used to energize a loudspeaker).

**IDEAL ELEMENTS** An ideal element *exhibits a specific property*, regardless of the frequency of operation or the impressed voltage and current flow through the element. For example, an ideal resistor of  $50\ \Omega$  remains  $50\ \Omega$  whether it operates at direct current or at hundreds of gigahertz (GHz) or has 1 or 10 000 V impressed across its terminals. Ideal elements, in terms of terminal voltage and current, are defined as follows:

**a. Resistor  $R$**  Current  $i$  (amperes) in an ideal resistor of  $R$  (ohms) is equal to the voltage  $e$  (volts) across  $R$  divided by  $R$ :

$$i = \frac{e}{R} \quad (6.1a)$$

Equation (6.1a) is the basic expression of Ohm's law. By algebraic manipulation, one also may write

## 6-2 Practical Circuit Analysis

$$e = iR \quad (6.1b)$$

$$R = \frac{e}{i} \quad (6.1c)$$

**b. Inductor  $L$**  The voltage  $e_L$  (volts) across an ideal inductor equals the value of the inductance  $L$  (henrys) multiplied by the rate of change of current with respect to time (amperes per second). A symbol used for the rate of change of current is  $\Delta i/\Delta t$ , where  $\Delta$  (*delta*) denotes a *small change*. Hence,

$$e_L = L \frac{\Delta i}{\Delta t} \quad (6.2)$$

Assume, for example, that at time  $t_1 = 6$  s, current  $i_1$  in the inductor is 1 A. At  $t_2 = 8$  s, the current is  $i_2 = 2$  A. Hence,  $\Delta i = 2 - 1 = 1$  A, and  $\Delta t = 8 - 6 = 2$  s. The rate of change of current is, therefore,  $\Delta i/\Delta t = \frac{1}{2}$  A/s.

**c. Capacitor  $C$**  The current  $i_C$  (amperes) in an ideal capacitor equals the value of capacitance  $C$  (farads) multiplied by the rate of change of the impressed voltage with respect to time (volts per second). The symbol for the rate of change of the impressed voltage is  $\Delta e/\Delta t$ . Hence,

$$i_C = C \frac{\Delta e}{\Delta t} \quad (6.3)$$

**d. Voltage source  $e$**  An ideal voltage source is a source whose output voltage is constant, regardless of the current it supplies to a circuit. A practical voltage source, with a low output resistance, approximates an ideal voltage source.

**e. Current source  $i$**  An ideal current source is a source whose output current is constant, regardless of the resultant voltage appearing across the circuit to which the source is connected. Symbols used for ideal voltage and current sources are given in Fig. 6.1.

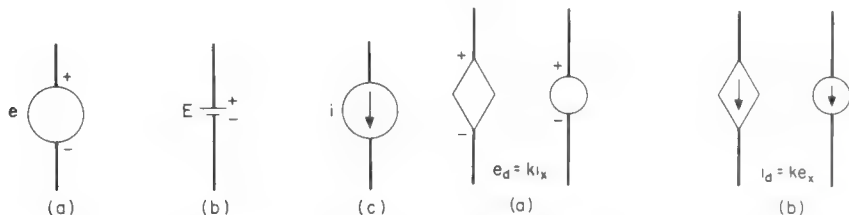
The symbol of Fig. 6.1a is general and can be used to denote a dc or time-varying voltage source. Figure 6.1b shows an ideal dc voltage source. The symbol for the ideal current source of Fig. 6.1c is used for dc and time-varying current sources.

**f. Dependent (controlled) sources** The sources in Fig. 6.1 are *independent* because their value does not depend on the circuit to which they are connected. In the analysis of transistor amplifiers, for example, the model of a transistor may contain a current or voltage source whose value *depends* on a specific current or a voltage elsewhere in the circuit. These are called *dependent*, or *controlled*, sources. Electrical symbols for controlled sources are provided in Fig. 6.2. Either the diamond or the circle symbol may be used.

In Fig. 6.2a, the value of the controlled voltage source  $e_d$  depends on current  $i_x$  flowing in another part of the circuit; this is an example of a *current-controlled (dependent) voltage source*. If  $e_d$  depended on a voltage  $e_x = ke_x$ , it would be called a *voltage-controlled voltage source*.

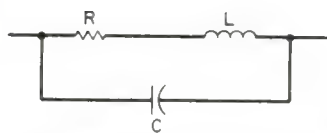
Refer to Fig. 6.2b. The value of the controlled current source  $i_d$  depends on voltage  $e_x$ ; this is an example of a *voltage-controlled current source*. If  $i_d$  depended on a current  $i_x = ki_x$ , it would be called a *current-controlled current source*.

**MODEL** A model is a representation of a *physical* device or circuit by *ideal elements*. Consider, for example, a resistor of  $R$  ohms. At very high frequencies, such as in the giga-

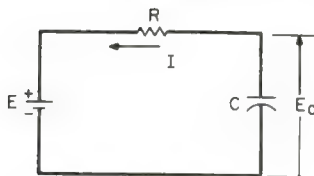


**Fig. 6.1** Symbols for ideal sources. (a) Dc or time-varying voltage source. (b) Dc voltage source. (c) Dc or time-varying current source.

**Fig. 6.2** Examples of controlled (dependent) sources. (a) Current-controlled voltage source. (b) Voltage-controlled current source.



**Fig. 6.3** A possible model of a resistor operating at very high frequencies.



**Fig. 6.4** Series RC circuit excited by a dc source.

hertz range, the lead inductance and shunt capacitance enter into the picture. Hence, a possible model of a resistor operating at high frequencies may appear as in Fig. 6.3. Note that the model is composed of ideal elements  $R$  and  $L$  in series, shunted by an ideal capacitor  $C$ .

**LINEAR ELEMENT** A linear element is an element in which a change of its input (such as voltage) results in a proportionate change in its response, or output (such as current). If, for example, the voltage across an  $8\text{-}\Omega$  resistor is  $16\text{ V}$ , then  $I_1 = \frac{16}{8} = 2\text{ A}$ . If the voltage is *doubled* to  $32\text{ V}$ , then  $I_2 = \frac{32}{8} = 4\text{ A}$ , and the current is also *doubled*.

**NONLINEAR ELEMENT** An element that is not linear is said to be nonlinear. In a nonlinear element, its response is not proportional to a change in its input. An example of a nonlinear element is a *thermistor*, whose resistance depends on its current flow.

**LINEAR CIRCUIT** A circuit that contains only linear elements is called linear. (The term *network* is also used for circuit.)

**NONLINEAR CIRCUIT** A circuit that contains one or more nonlinear elements is nonlinear.

**SERIES CIRCUIT** A circuit in which the current in each element is the same is a series circuit.

**PARALLEL CIRCUIT** A circuit in which the voltage across each element is the same is a parallel circuit.

## 6.4 RC and RL CIRCUITS

Consider the series RC circuit of Fig. 6.4. Initially, that is, at time equals zero ( $t = 0$ ), capacitor  $C$  acts as a *short circuit*. Hence, the current  $I$  at  $t = 0$ , denoted by  $I(0)$ , is

$$I(0) = \frac{E}{R} \quad (6.4)$$

The product  $RC$  is the *time constant*  $T$  (seconds) of the circuit,

$$T = RC \quad (6.5)$$

where  $R$  is in ohms and  $C$  in farads. After an elapsed time of approximately 5 time constants ( $5T$ ), the circuit is said to be in *steady state*, denoted by  $t = \infty$ . In steady state, the voltages and currents in the circuit are not changing and have settled down to a fixed value. For dc excitation in steady state,  $\Delta e/\Delta t = 0$ . From Eq. (6.3), if  $\Delta e/\Delta t = 0$ , the current in the capacitor is also zero ( $i_C = 0$ ). Therefore, a capacitor acts as an *open circuit* to direct current in the steady state.

**example 6.1** In Fig. 6.4 assume that  $E = 100\text{ V}$ ,  $R = 1\text{ k}\Omega$ , and  $C = 1\text{ }\mu\text{F}$ . Determine (a) time constant  $T$ , (b) initial and steady-state values of current, and (c) initial and steady-state values of voltage across the capacitor.

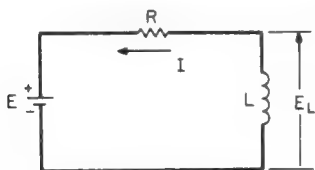
**solution** (a) From Eq. (6.5),  $T = RC = 10^3 \times 10^{-6} = 10^{-3}\text{ s}$  ( $1\text{ ms}$ ).  
 (b) Because at  $t = 0$  the capacitor behaves as a short circuit to direct current,  $I(0) = E/R = 100/1000 = 0.1\text{ A}$ . In steady state the capacitor acts as an open circuit to direct current, and  $I(\infty) = 0$ .

## 6-4 Practical Circuit Analysis

(c) Because  $C$  acts as a short circuit at  $t = 0$  (and has no charge),  $E_C(0) = 0$  V. In steady state, no current flows in  $C$ , and the capacitor is fully charged by the source voltage; therefore,  $E(\infty) = 100$  V.

Consider now the series  $RL$  circuit of Fig. 6.5. To direct current,  $L$  acts as an open circuit at  $t = 0$  and as a short circuit in steady state ( $t = \infty$ ). The current in the circuit is, therefore, maximum in steady state for dc excitation.

The time constant  $T$  (seconds) is



**Fig. 6.5** Series  $RL$  circuit excited by a dc source.

$$T = \frac{L}{R} \quad (6.6)$$

where  $L$  is in henrys and  $R$  in ohms. At  $t = 0$ , the voltage across the inductor is  $E_L(0) = E$  volts. In steady state,  $\Delta i / \Delta t = 0$ ; hence, from Eq. (6.2),  $E_L(\infty) = 0$  V.

**example 6.2** In Fig. 6.5, let  $E = 25$  V,  $R = 100 \Omega$ , and  $L = 1$  H. Determine (a)  $T$ , (b)  $I(0)$  and  $I(\infty)$ , (c)  $E_L(0)$  and  $E_L(\infty)$ .

**solution** (a) From Eq. (6.6),  $T = L/R = 1/100 = 0.01$  s.

(b) At  $t = 0$ ,  $L$  acts as an open circuit to direct current; hence,  $I(0) = 0$ . In steady state,  $L$  acts as a short circuit to direct current, and  $I(\infty) = E/R = 25/100 = 0.25$  A.

(c)  $E_L(0) = 25$  V and  $E_L(\infty) = 0$  V.

For times between approximately 0 and  $5T$ , the current  $i_C$  and voltage  $e_C$  as a function of time for a series  $RC$  circuit excited by a dc source (Fig. 6.4) are

$$i_C = \frac{E}{R} \epsilon^{-t/(RC)} \quad (6.7)$$

and

$$e_C = E(1 - \epsilon^{-t/(RC)}) \quad (6.8)$$

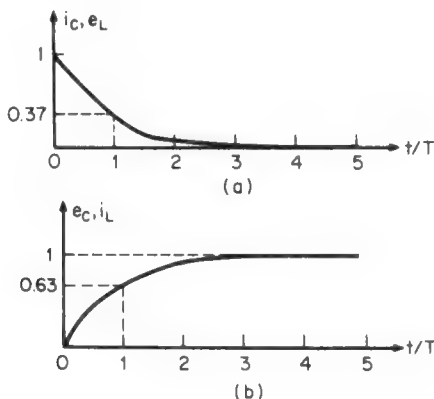
For the series  $RL$  circuit excited by a dc source (Fig. 6.5),

$$i_L = \frac{E}{R} (1 - \epsilon^{-tR/L}) \quad (6.9)$$

and

$$e_L = E \epsilon^{-tR/L} \quad (6.10)$$

for times between 0 and  $5T$ .



**Fig. 6.6** Performance of series  $RC$  and  $RL$  circuits energized by a dc voltage source: (a)  $i_C$  and  $e_L$  as a function of time. (b)  $e_C$  and  $i_L$  as a function of time.

Equations (6.7),  $i_C$ , and (6.10),  $e_L$ , are of the same form and so are plotted in Fig. 6.6a. For convenience, the maximum value of either  $i_C$  or  $e_L$  is denoted by 1. (This is called *normalizing* the maximum value to 1.) The unit for the  $x$  axis is  $t/T$ . Note that when  $t$  is equal to five time constants ( $5T$ ), the current  $i_C$  in a capacitor and the voltage  $e_L$  across an inductor are approximately equal to zero. In a single time constant ( $t = T$ ),  $i_C$  (or  $e_L$ ) falls to approximately 0.37, or 37 percent of its maximum value.

Figure 6.6b is a plot of  $e_C$  and  $i_L$  [Eqs. (6.8) and (6.9)]. The maximum value is again denoted by 1, and in  $t = 5T$  steady state has been reached. For  $t = T$ ,  $e_C$  (or  $i_L$ ) has reached 63 percent of the maximum value.

## 6.5 ENERGY AND POWER

Energy is the ability to do work. In SI units, the unit of energy is the *joule* (J); 1 joule equals 1 *watt-second* (W·s). Power is the *rate of expending energy*; its unit is *joule per second* which equals 1 *watt*. Energy  $w$  equals the product of power  $p$  and time  $t$ :

$$w = pt \quad (6.11)$$

For electric circuits, power  $p$  may be expressed by

$$p = ei \quad (6.12a)$$

$$= \frac{e^2}{R} \quad (6.12b)$$

$$= i^2 R \quad (6.12c)$$

where  $R$  is the resistance,  $e$  the impressed voltage across the resistance, and  $i$  the current flowing in the resistance.

Only *resistors* are capable of *dissipating energy*; capacitors and inductors only *store energy*. The energy stored in a capacitor  $w_C$  is

$$w_C = \frac{Ce^2}{2} \quad (6.13)$$

where  $e$  is the voltage across  $C$ . For an inductor the stored energy  $w_L$  is

$$w_L = \frac{Li^2}{2} \quad (6.14)$$

where  $i$  is the current flowing in the inductor.

**example 6.3** Referring to Example 6.1, determine the energy stored in the capacitor in steady state.

**solution** In Example 6.1,  $C = 10^{-6}$  F, and the voltage across  $C$  equals 100 V. From Eq. (6.13),  $w_C = Ce^2/2 = 10^{-6} \times 100^2/2 = 5 \times 10^{-3}$  J.

**example 6.4** Referring to Example 6.2, determine (a) the power dissipated in  $R$  and (b) the energy stored in  $L$ , in steady state.

**solution** (a) In Example 6.2,  $I = 0.25$  A in steady state;  $L = 1$  H; and  $R = 100 \Omega$ . From Eq. (6.12c), by letting  $i = I$ ,  $p = I^2 R = (0.25)^2 \times 100 = 6.25$  W.

(b) From Eq. (6.14),  $w_L = LI^2/2 = 1 \times (0.25)^2/2 = 31.75 \times 10^{-3}$  J.

## 6.6 CIRCUIT LAWS AND THEOREMS

A number of simple, yet powerful, laws and theorems for the analysis of circuits are available to the technician. Although the application of these laws and theorems is demonstrated first in the analysis of dc circuits, they are also applicable in the analysis of ac circuits. For dc circuits the quantities of concern, such as voltage and current, are always expressed by real numbers. For ac circuits, however, these quantities are generally represented by complex numbers. Complex numbers and the application of the circuit laws and theorems to ac circuits are reviewed later in the chapter.

**KIRCHHOFF'S LAWS** Kirchhoff's voltage law (KVL) and current law (KCL) provide the technician with two useful tools for the analysis of circuits, regardless of their complexity. The voltage law states that the algebraic sum of voltages in a closed-circuit path is always

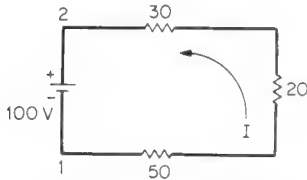
## 6-6 Practical Circuit Analysis

equal to zero. The current law states that the algebraic sum of currents leaving a circuit node is always equal to zero. A circuit node may be thought of as a connecting point that is common to two or more circuit elements.

In the application of KVL and KCL, it is essential that we designate the polarities of voltages across elements and the direction of electron flow at a circuit node. In this chapter we adopt the following conventions:

1. In traversing a closed-circuit path, a *plus sign* precedes a voltage quantity, such as  $e$  or  $iR$ , in going *with* the direction of electron flow or from a *low* to a *high* potential. In going *against* the electron flow or from a *high* to a *low* potential, a *minus sign* precedes the voltage quantity.
2. Current flowing *away* from a circuit node will have a *negative sign*; current flowing *toward* a circuit node will have a *positive sign*. A current source pointing to a node will have a *negative sign*.

The application of KVL and KCL is illustrated in the next group of examples.



**Fig. 6.7** Finding current  $I$  in a series circuit by KVL (Example 6.5).

**example 6.5** Find current  $I$  in Fig. 6.7.

**solution** Starting at point 2 and traversing the circuit in the direction (counterclockwise) of current  $I$ , we have, according to KVL and our conventions,

$$-100 + 30I + 20I + 50I = 0$$

From point 2 to point 1, we go from a *high* to a *low* potential, hence the minus sign before 100. In traversing the remainder of the circuit, we go in the direction of electron flow. The voltage across each resistor, therefore, has a positive sign.

Bringing the  $-100$  term to the right side of the equation and simplifying, we have

$$(30 + 20 + 50)I = 100$$

or

$$100I = 100$$

Solving yields  $I = \frac{100}{100} = 1$  A.

**example 6.6** Figure 6.8 shows a circuit having two closed-circuit paths. This configuration is referred to as a *two-mesh*, or *two-loop*, circuit. Circuits containing two or more meshes are fairly common.

In Fig. 6.8, currents  $I_1$  and  $I_2$  are assumed to be flowing counterclockwise. (If it turns out that the actual current flows in a direction opposite to that assumed, the solved current will have a minus sign.) Applying KVL, we write the necessary equations and solve for the currents.

**solution** Starting at point 1 for the first mesh, we obtain

$$-100 + (2 + 6)I_1 - 6I_2 = 0$$

or

$$100 = 8I_1 - 6I_2$$

Dividing by 2 yields

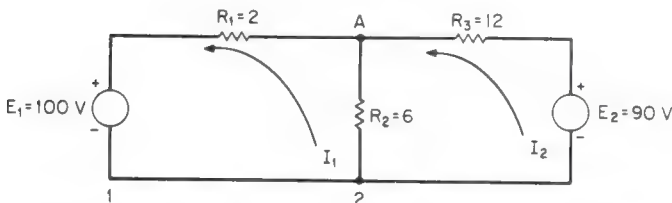
$$50 = 4I_1 - 3I_2 \quad (a)$$

Starting at point 2 in the second mesh, we have

$$-6I_1 + (12 + 6)I_2 + 90 = 0$$

or

$$-90 = -6I_1 + 18I_2$$



**Fig. 6.8** Finding currents  $I_1$  and  $I_2$  in a two-mesh circuit by KVL (Example 6.6).



Thus

$$-45 = -3I_1 + 9I_2 \quad (b)$$

Equations (a) and (b) constitute a pair of linear simultaneous equations. One may use determinants or substitution to find the currents. With the latter procedure and by solving Eq. (b) for  $I_2$ , we have

$$I_2 = \frac{-45 + 3I_1}{9} = -5 + \frac{I_1}{3} \quad (c)$$

Substitution of Eq. (c) in Eq. (a) yields

$$\begin{aligned} 50 &= 4I_1 - 3\left(-5 + \frac{I_1}{3}\right) \\ &= 4I_1 + 15 - I_1 = 15 + 3I_1 \end{aligned}$$

Solving gives  $I_1 = (50 - 15)/3 = 11.67$  A. Substituting  $I_1 = 11.67$  A in Eq. (b) gives

$$-45 = -3(11.67) + 9I_2 = -35 + 9I_2$$

Solving yields  $I_2 = -1.11$  A. The minus sign denotes that the actual direction of current flow in the second mesh is *clockwise*.

In the literature, the sum of resistances in a mesh is referred to as the *self-resistance* of the mesh. The resistance *common* to two meshes is called the *mutual resistance*. To return to Fig. 6.8, the sum  $R_1 + R_2$  is the self-resistance of mesh I and is denoted by  $R_{11} = R_1 + R_2$ . Resistors  $R_2 + R_3$  constitute the self-resistance of mesh II ( $R_{22} = R_2 + R_3$ ). Resistor  $R_2$  is the mutual resistance linking meshes I and II ( $R_{12} = R_{21} = R_2$ ).

**example 6.7** Apply KCL and solve for  $E_1$  and  $E_2$  in Fig. 6.9.

**solution** There are two independent nodes: 1 and 2. Node  $N$  is a *common*, or *reference*, node for  $E_1$  and  $E_2$ . Taking the direction of currents as flowing from right to left and from bottom to top, according to our conventions, at node 1 we have

$$-\frac{10 - E_1}{1} + \frac{E_1}{2} + \frac{E_1 - E_2}{1} = 0 \quad (a)$$

The current flowing in a resistor is equal to the voltage difference across the resistor divided by its value. Thus, the current in  $R_1$  is  $(10 - E_1)/1$ ; in  $R_2$ , it is  $E_1/2$ ; etc.

Simplifying Eq. (a), we get

$$\begin{aligned} 10 &= E_1(1 + \frac{1}{2} + 1) - \frac{E_2}{1} \\ &= 2.5E_1 - E_2 \end{aligned} \quad (b)$$

At node 2,

$$-\frac{E_1 - E_2}{1} + \frac{E_2}{2} - 10 = 0 \quad (c)$$

or

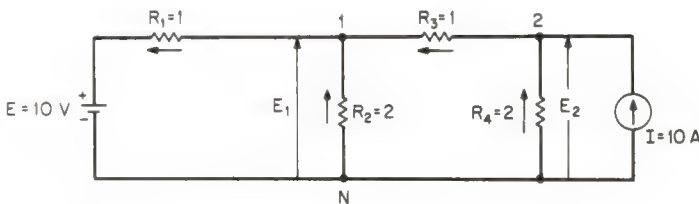
$$10 = -E_1 + 1.5E_2 \quad (d)$$

Solving Eq. (c) for  $E_1$  gives

$$E_1 = -10 + 1.5E_2 \quad (d)$$

Substituting Eq. (d) in Eq. (b) yields

$$\begin{aligned} 10 &= 2.5(-10 + 1.5E_2) - E_2 \\ &= -25 + 2.75E_2 \end{aligned}$$



**Fig. 6.9** Finding voltages  $E_1$  and  $E_2$  by KCL (Example 6.7).



## 6-8 Practical Circuit Analysis

Hence,

$$E_2 = \frac{10 + 25}{2.75} = 12.7 \text{ V}$$

and

$$E_1 = -10 + 1.5(12.7) = 9.05 \text{ V}$$

**CONDUCTANCE** Conductance  $G$  is equal to 1 divided by the resistance  $R$ :

$$G = \frac{1}{R} \quad (6.15)$$

The unit for conductance is the *stemen* (SI).

By letting  $G_1 = 1/R_1$ ,  $G_2 = 1/R_2$ ,  $G_3 = 1/R_3$ , and  $G_4 = 1/R_4$ , Eqs. (b) and (c) in Example 6.7 may be expressed by

$$\begin{aligned} 10 &= (G_1 + G_2 + G_3)E_1 - G_3E_2 \\ 10 &= -G_3E_1 + (G_3 + G_4)E_2 \end{aligned}$$

where  $G_1 + G_2 + G_3 = G_{11} = 1 + 0.5 + 1 = 2.5 \text{ S}$  is the *self-conductance* of node 1,  $G_3 = G_{12} = G_{21} = 1 \text{ S}$  is the *mutual conductance* between nodes 1 and 2, and  $G_3 + G_4 = G_{22} = 1 + 0.5 = 1.5 \text{ S}$  is the *self-conductance* of node 2.

# Chapter 7

## Meters and Measurements

### 7.1 INTRODUCTION

It is important for a technician to have an understanding of the difficulties involved in making exact measurements if he is to accurately evaluate the readings on his meter. The first part of this chapter is devoted to the problems of obtaining and interpreting accurate measurements. This information is also valuable for someone who is planning to purchase meters.

Measurements of current are made indirectly—either by the effect of their magnetic field or by the heat produced when the current flows through a resistance. A *meter movement* is the electromechanical part of a meter that operates in accordance with either the magnetic or heat effect of the current. A number of different meter movements are discussed in this chapter.

By using a sensitive meter movement for the indicator, it is possible to make ammeters, voltmeters, ohmmeters, and electronic multimeters for a wide variety of electrical measurements. The required circuitry is discussed in this chapter, including the procedure for calculating the shunt and series resistance values for the meters.

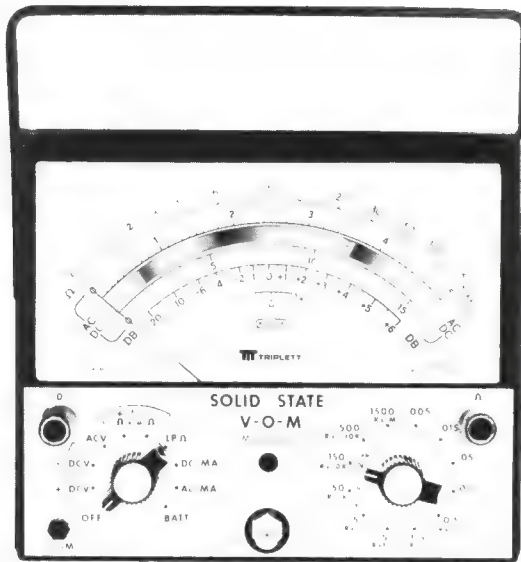
When accurate measurements of resistance, inductance, or capacitance are to be made, a bridge circuit is generally used. The theory of bridge circuits and the circuits for various bridges are included in this chapter.

### 7.2 PARALLAX AND OTHER PROBLEMS IN USING INSTRUMENTS

Consider the problem of reading an ammeter. If two different people sight along the needle of the meter movement, the slight difference in the position of their eyes while they are making the reading will make a difference in their interpretation of the current value. You may have encountered the same problem when reading a speedometer in an automobile. The driver, who sits *behind* the speedometer, will see the speed as one value, whereas the passenger looking at the speedometer from his position will interpret its reading as being entirely different. This problem is known as *parallax*. It can be reduced somewhat by placing a mirror behind the meter.

Figure 7.1 shows a meter with a mirrored scale. The proper way to use this meter is to sight along the needle with one eye in such a way that you cannot see the reflection of the needle in the mirror. When you do this, your eye is looking exactly perpendicular to the needle, and your measurements will be most accurate.

Even though you eliminate the problem of parallax, you still have the problem of interpreting the position of the meter when it falls between marks on the scale. As with the yardstick, one person will see a reading as six-tenths of the way between the marks whereas another person will see it as seven-tenths of the way between the marks. The interpretation of the reading is very important. To improve accuracy in



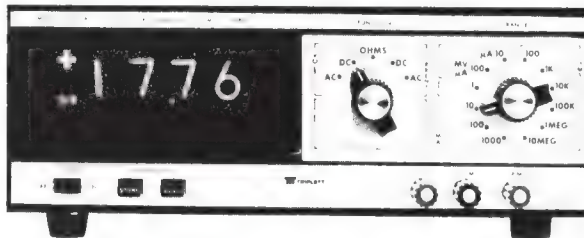
**Fig. 7.1** A meter with a mirrored scale. (Courtesy Triplett Electrical Instrument Company)

this regard some practice will be required. This is one of the advantages of a *digital voltmeter* like the one shown in Fig. 7.2. Instead of using a meter movement with a needle, this type of meter actually shows the digits of the reading accurately to four or five decimal places. This eliminates the problem of parallax.

The most difficult part of making a reading with a meter is calibration. Here we encounter the same problem that we had with the yardstick. Two different voltmeters will take the same voltage measurements and give two different readings. Which one is correct? The answer is that probably neither one is 100 percent correct. Again, if we wanted a very, very, precise measurement, we would need to use a more expensive meter, and it would have to be accurately calibrated at the Bureau of Standards.

Another problem in making accurate instrument measurements is that when the instrument is inserted into the circuit, changes in circuit current, frequency, voltage, and resistance may occur.

When all these factors are taken into consideration, we are faced with a difficulty in making highly accurate measurements. If a vacuum tube is supposed to have 100 V on its plate, it will operate satisfactorily if the voltage is 101 or 99 V. In fact, if the voltmeter reading at the plate indicates that the voltage is within 5 or even 10 percent of the manufacturer's stated value, we can usually presume that the stage is operating satis-



**Fig. 7.2** A digital, solid-state volt-ohm-milliammeter. (Courtesy Triplett Electrical Instrument Company)

factorily. The same is true of transistor circuits. When we make a voltage, current, or resistor check during a troubleshooting procedure, we must learn to interpret the reading we get in terms of the circuit complexity and decide whether the inaccuracy is sufficient to cause the trouble we are looking for. Learning to read the instruments is the easiest part. Learning to interpret the readings is more difficult.

### 7.3 PERCENT ACCURACY

When the percent accuracy of an instrument is stated by a manufacturer, the value given usually refers to the accuracy at the *full-scale reading*. As an example, suppose that the accuracy of a 0- to 10-mA meter movement is given as  $\pm 10$  percent. This means that a full-scale reading of 10 mA may actually be due to a current that is

$$10 \text{ mA} + 10\% \text{ of } 10 \text{ mA} = 11 \text{ mA}$$

or

$$10 \text{ mA} - 10\% \text{ of } 10 \text{ mA} = 9 \text{ mA}$$

The manufacturer is saying, in so many words, that his instrument may be in error by as much as 1 mA for the full-scale reading. To show why this is misleading (unless we fully understand the rating), suppose that the meter in the above example is used to measure a current of 5 mA, and the reading is in error by 1 mA—that is, the meter indicates 6 mA when it is actually measuring a current of 5 mA. Now the percent error is much greater than the  $\pm 10$  percent stated for the meter at a full-scale reading. The percent error is calculated as follows:

$$\% \text{ error} = \frac{\left( \begin{array}{c} \text{current indicated} \\ \text{by meter} \end{array} \right) - \left( \begin{array}{c} \text{actual current} \\ \text{being measured} \end{array} \right)}{\text{current indicated by meter}} \quad (7.1)$$

By using the *indicated* value of current of 6 mA for an *actual* current flow of 5 mA, the percent error is calculated as follows:

$$\begin{aligned} \% \text{ error} &= \frac{\left( \begin{array}{c} \text{current indicated} \\ \text{by meter} \end{array} \right) - \left( \begin{array}{c} \text{actual current} \\ \text{being measured} \end{array} \right)}{\text{current indicated by meter}} \quad (7.1) \\ \% \text{ error} &= \frac{6 - 5 \text{ mA}}{6 \text{ mA}} = 16.1\% \end{aligned}$$

Therefore, an error of 1 mA will cause an error of 10 percent at full scale (which is the manufacturer's rating) but will cause an error of 16.6 percent at half scale. The manufacturer's rating, then, is for the *maximum possible accuracy* when the meter is deflected full scale.

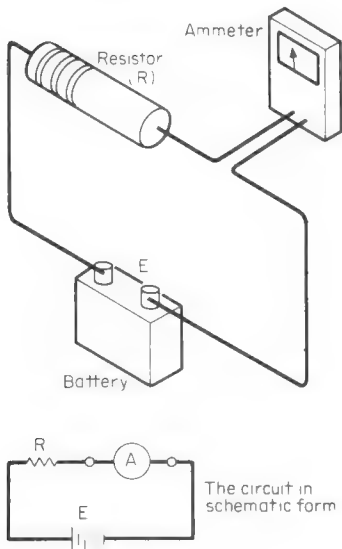
It might be argued that an error of 1 mA at full scale is the maximum possible error of the instrument, and that the error when reading at half scale will be correspondingly less. However, this is not necessarily true. In fact, the reading error at half scale *may* actually be *greater* than at full scale. For this reason the calibration of an instrument should be checked (whenever possible) in the part of the scale where most of the readings are going to take place.

### 7.4 BASIC INSTRUMENTS AND THEIR USE FOR SERVICING ELECTRONICS EQUIPMENT

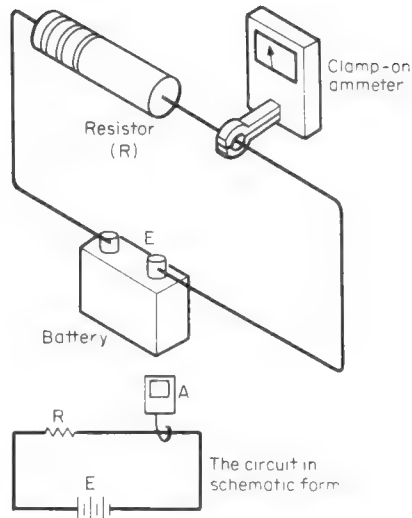
Ammeters, voltmeters, and ohmmeters are the three basic instruments used for measurements in electric and electronic circuits.

**AMMETERS** The instrument used for measuring electric current is an *ammeter*. See Fig. 7.3. (When designed for measuring very small currents, it is called a milliammeter or a microammeter.) In order to measure a current flow with an ordinary ammeter, it is necessary to have the current flow *through* the instrument, as shown in Fig. 7.3a. A special instrument, called a clamp-on ammeter, operates by measuring the

## 7-4 Meters and Measurements



**Fig. 7.3a** To measure current with an ordinary ammeter it is necessary to open the circuit and insert the ammeter so that the current being measured flows *through* it.



**Fig. 7.3b** The clamp-on ammeter can measure current without the need of opening the circuit to insert the instrument.

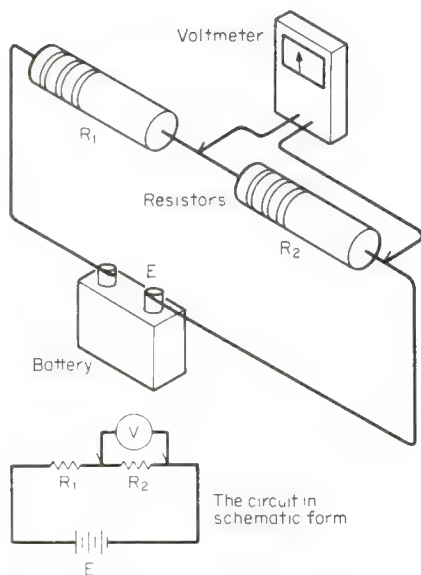
**Fig. 7.3** Two instruments used for measuring current.

strength of the magnetic field around the current-carrying wire, as shown in Fig. 7.3b. The clamp-on ammeter is more convenient for general servicing because it does not require that the circuit be opened in order to insert the ammeter in series with the circuit components. Clamp-on ammeters are a relatively recent introduction in the instrument field, and they are becoming more and more popular with electronic servicemen.

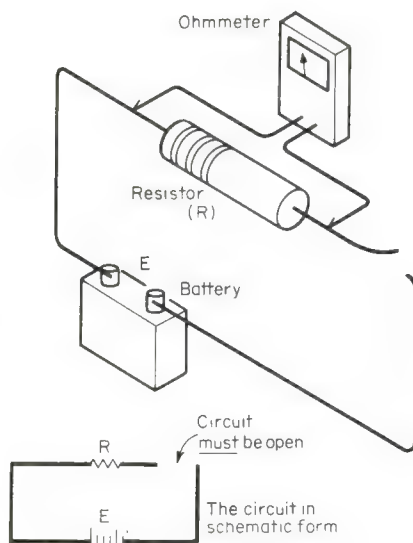
It is seldom necessary to measure the current in vacuum-tube receivers. It is usually easier to measure the voltage across a resistor, and if it is desired to find the current through that resistor, it is a simple matter to apply Ohm's law ( $I = E/R$ ). Vacuum tubes are voltage-operated devices, and therefore we are interested in the plate and grid voltages when troubleshooting this type of receiver. In transistor receivers, however, current measurements are sometimes important. Specifically, we may be concerned about the total amount of current drawn by a transistor radio because this is an indication of whether the transistor amplifiers are operating properly.

**VOLTMETERS** Manufacturer's literature often gives values of voltages that should be measured at various points in a receiver. In order to make these measurements, a voltmeter is used. To measure a voltage, the voltmeter must be placed *across* the circuit or component (not in series with it). Figure 7.4 shows how a voltmeter is employed to measure the voltage across a resistor.

Generally the voltage measurements in a receiver are taken with respect to a *common point*, often called the *ground point*. This means that one of the meter probes is connected to the ground point, and the other probe is connected to the point where the voltage is being measured. If the meter reads +15 V, it means that the point in question (that is, the point where the voltage is being measured) is 15 V positive with respect to ground. There are exceptions, however. If we want to measure the actual grid voltage of a vacuum tube, we must connect the voltmeter between the grid and cathode because, by definition, the grid voltage is the voltage on the grid *with respect to the cathode* (not with respect to ground).



**Fig. 7.4** To measure a voltage the meter probes are placed across the difference of potential.



**Fig. 7.5** To measure a resistance the ohmmeter is placed across the resistor. The circuit must be opened so there will be no voltage across the resistor.

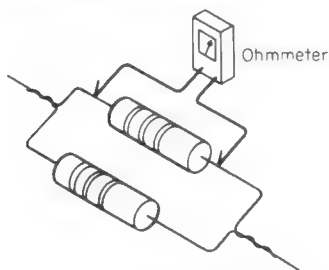
Many voltmeters are not sensitive enough to measure the difference in voltage between the emitter and the base of a properly operating transistor amplifier. However, the base or emitter voltage is sometimes measured with respect to ground for the purpose of troubleshooting.

**OHMMETERS** An ohmmeter is an instrument used to measure resistance. Most ohmmeters operate by sending a current through the unknown resistance and measuring the resulting voltage drop across that resistor. A voltmeter, having a scale that is actually marked in ohms, is used for the indicator. In using an ohmmeter it is very important that it be in a circuit that has *no voltage applied*—that is, no voltage other than the voltage source inside the meter. To make a resistance measurement, the ohmmeter is placed *across* the component to be measured, as shown in Fig. 7.5.

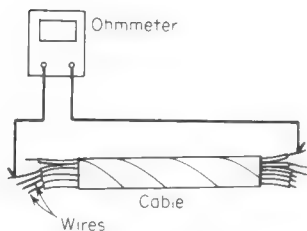
In using an ohmmeter it is always important to be sure that the component being measured is not in parallel with another component. If the components are in parallel, as shown in Fig. 7.6, the ohmmeter will measure the equivalent resistance of the parallel combination. If both resistors in this circuit are  $1000\ \Omega$  in value, the equivalent resistance of the parallel combination is  $500\ \Omega$ ; and this is the resistance that will be indicated by the ohmmeter. It will appear, then, that the  $1000\text{-}\Omega$  resistor is defective because it measures only half of its rated value. Parallel branches of resistance are not always so obvious as the one shown in Fig. 7.6. It may be necessary to check the circuit schematic diagram to make sure that there is not a parallel path. If there is a parallel path, then it is necessary to disconnect the resistor before measuring its ohmic value.

Ohmmeters are also convenient instruments for use in measuring *continuity*. Continuity simply means that there is a continuous path for current to flow from one point to another. As an example of a continuity measurement, consider the multiwire cable of Fig. 7.7. There are five wires coming out of the cable at each end. It is desired to trace one wire through the cable. To do this, one probe of the meter is connected to one of the wires at one side of the cable. Then, the other probe is touched to the wires (one at a time) at the other side until the ohmmeter reads zero resistance. The wire that indicates zero resistance is obviously connected to the other ohmmeter probe.

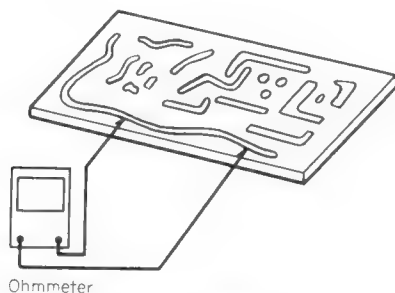




**Fig. 7.6** When making a resistance measurement, the component being measured must not be in parallel with another resistance path.



**Fig. 7.7** An ohmmeter can be used for continuity checks.



**Fig. 7.8** Using an ohmmeter to check a printed-circuit board for cracks in the metal foils.

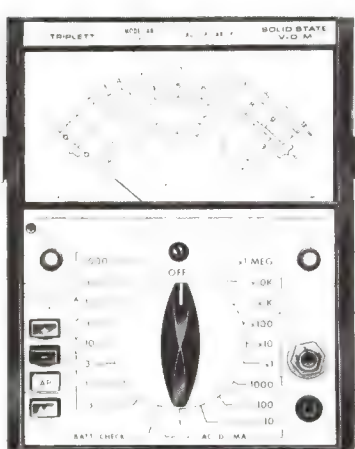
Continuity measurements are a valuable aid in servicing printed-circuit boards. If there is a tiny crack in a conductor on the board, it may not be visible to the naked eye. However, by placing an ohmmeter across the conductor and flexing the board, the ohmmeter will show that the path is being intermittently opened and closed. This type of test is shown in Fig. 7.8.

**MULTIMETERS** Instead of having separate instruments for measuring current, voltage, and resistance, all three may be included in a single instrument, as shown in Fig. 7.9a. Figure 7.9a presents an example of a volt-ohm-milliammeter (VOM). It can measure current, voltage, and resistance and has high input impedance.

The transistorized volt-ohmmeter of Fig. 7.9b has a very high input impedance (usually about 20 M $\Omega$ ). The solid-state equivalent of a VTVM (vacuum-tube voltmeter), is the FET volt-meter shown in Fig. 7.9c. This instrument is battery-operated, and therefore it has the advantage of portability over the VTVM. Like the VTVM, it has the advantage of a very high input impedance. Both the VTVM and FET voltmeters have a higher input impedance than the VOM. They can measure voltages without noticeably changing the circuit voltage. Figure 7.10 shows why this is important. The 5-k $\Omega$  resistor of Fig. 7.10a has two volts across it. When the VTVM is placed across it, we have 5 k $\Omega$  and 20 M $\Omega$  in parallel. For all practical purposes, the voltmeter is an open circuit, and only 5 k $\Omega$  is in the circuit during measurement.

Figure 7.10b shows what happens when the voltage across the 5-k $\Omega$  resistor is measured with a voltmeter that has only 5000  $\Omega$  of resistance. The two resistances in parallel have a combined resistance of 2.5 k $\Omega$ . Since the circuit resistance is decreased by the presence of the voltmeter, both the voltage across the resistor and the circuit current are changed during measurement. In general, the higher the input resistance of the measuring device, the less its effect on the circuit being measured, and the more accurate the measurement.





(a)



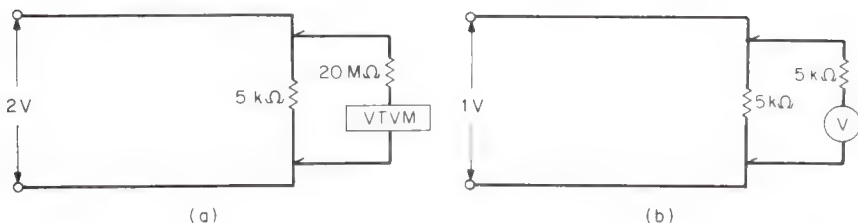
(b)



(c)

**Fig. 7.9** Three examples of multimeters: (a) Volt-ohm-milliammeter. (Courtesy Triplet Corporation) (b) Transistorized volt-ohmmeter. (Courtesy Triplet Corporation) (c) FET meter. (Courtesy Sencore, Inc.)

## 7-8 Meters and Measurements



**Fig. 7.10** This shows the importance of a high-impedance voltmeter: (a) The VTVM does not disturb the circuit when making measurements. (b) The low-impedance voltmeter reduces the circuit voltage.

## 7.5 METER MOVEMENTS IN MEASURING INSTRUMENTS

An ampere may be defined as a flow of electric current equal to one coulomb per second. In other words, an ampere of current is flowing whenever 6 250 000 000 000 000.0 electrons (that is:  $6.25 \times 10^{18}$  electrons) flow past a point on a conductor every second. Obviously, it would not be possible to count this number of electrons directly even if they could be seen. Therefore, some form of indirect measurement must be used to determine the amount of current flow. By "indirect method of measurement," we mean that the current is measured according to some effect that it produces in a circuit.

When an electric current flows through a circuit, there are several effects that are always produced. For example, heat is produced when current flows through a resistance. The amount of heat is directly related to: (a) the square of the amount of current flow, and (b) the amount of resistance in the circuit. This is evident from the equation for power dissipated by a resistor given in equation form as

$$P = I^2 R \quad (7.2)$$

where  $P$  = power dissipated in the form of heat, W

$I$  = current flow, A

$R$  = resistance,  $\Omega$

If the amount of heat produced in a resistor can be accurately measured, and if, at the same time, the value of resistance is known, then the current can be determined by transposing the above equation:

$$I = \sqrt{\frac{P}{R}} \quad (7.2a)$$

We cannot directly substitute the temperature for  $P$  in this equation. Temperature and power dissipation are two different things, and they must be treated as such. Thus, if we measure the amount of temperature rise due to a current flowing in a resistor, this temperature rise must be converted to the power dissipated in watts. The method of measuring power dissipation by the amount of heat generated is called the *calorimetric method*. Once the power is known, the current can be obtained from Eq. (7.2a).

The calorimetric method may seem like a rather roundabout method of measuring current, but it is very useful in certain applications. For example, a high-powered transmitter may be radiating 50 000 or even 100 000 W of power. This amount of power is difficult to measure accurately with ordinary measuring instruments, but the calorimetric method can be used. The output of the transmitter is fed through a *dummy load* (an artificial load) which is immersed in water. The temperature rise of the water is accurately measured, and from this information the output power can be calculated quite accurately. The manufacturer of equipment for making calorimetric measurements usually supplies charts for interpreting temperature rise into power.

The calorimetric method of measuring current can be used in almost any application where it is necessary to measure a current, specifically a large current, by precise measurements. However, it is much too inconvenient for such applications as servic-

ing electronic equipment. (Later we will discuss another method of measuring current by utilizing the amount of heat produced by the current.)

In addition to the heat generated by a current, there is also a magnetic field produced around the current-carrying conductor. The strength of this magnetic field is related to the amount of current flow. One way to measure the strength of a magnetic field around a current-carrying conductor is to measure the effects of a force produced when that magnetic field is inserted into another magnetic field with a known intensity. The force produced on a current-carrying conductor when placed in a magnetic field is given by Ampere's law. Stated in words, Ampere's law tells us that when a conductor carries an electric current in a magnetic field, there is a force on the conductor that is directly proportional to the current, directly proportional to the length of the conductor in the magnetic field, and directly proportional to the magnetic flux density of the field. Mathematically this law is stated as

$$F = \frac{BIL \sin \Theta}{10} \quad (7.3)$$

where  $F$  = force on conductor, dyn

$I$  = current flowing through conductor, A

$L$  = length of conductor, cm

$B$  = flux density of magnetic field into which conductor is inserted, G

$\Theta$  = angle between direction of current flow and direction of flux

Ampere's law is the principle upon which a number of measuring instruments are designed. In fact, most of the meter movements used in voltmeters, ammeters, ohmmeters, and multimeters are based on the principle stated by Ampere's law. The needle of the meter movement is moved upscale as a result of the magnetic field of a current flowing through a coil which reacts with either a permanent magnetic field or the field of another coil.

**METERS THAT UTILIZE THE MAGNETIC FIELD OF A CURRENT FOR MEASUREMENT** To summarize, instruments for measuring current employ indirect measurements based on either of two principles: the effect of heat produced by the current, or the effect of the magnetic field produced by the current. With the knowledge of these basic principles we can now study meter movements and see how they operate.

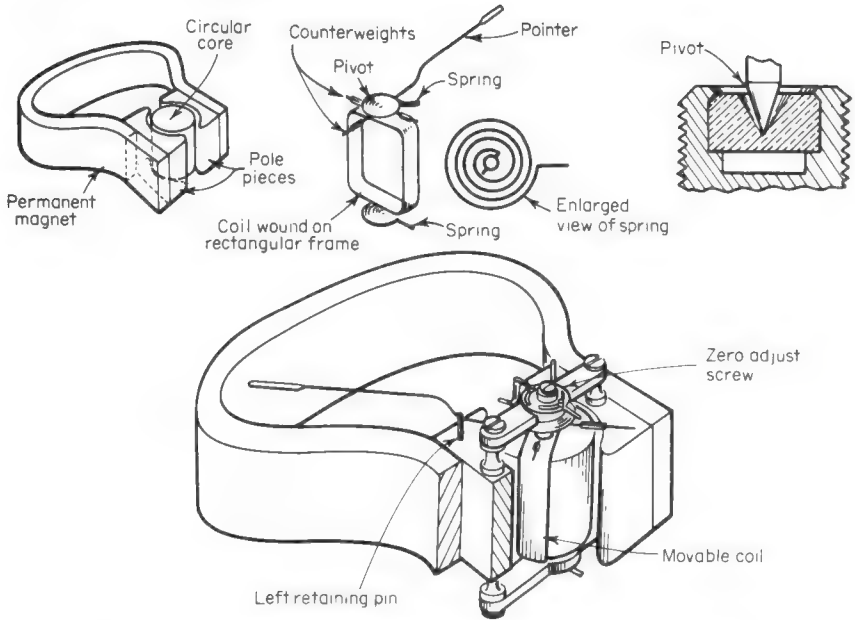
First, we will discuss measurements based on the use of a magnetic field accompanying a current; then we will discuss measurements based on the heat generated by a current.

**MOVING-COIL METERS** Instead of holding the coil stationary in a meter and having a piece of soft iron move in response to a magnetic field produced by a coil, some instruments are made in which a current-carrying coil itself is moved in a permanent-magnet field. Figure 7.11 shows an example of a moving-coil meter. In this case a movable coil is inserted into a permanent magnetic field. As current flows through the coil, the magnetic field surrounding the coil reacts with the permanent magnetic field and causes the coil to turn. The coil is connected in such a way that its magnetic field will always cause the pointer to move upscale. This is an important characteristic of this type of meter: in order to get an upscale movement of the pointer, it is necessary that the current flow through the coil in a specified direction! If the direction of current is reversed, the coil will try to move in the reverse direction. Very often this will destroy the meter movement. In order to measure an alternating current with this type of movement it is necessary to employ some type of a rectifying device to convert the alternating current to direct current flowing in the proper direction.

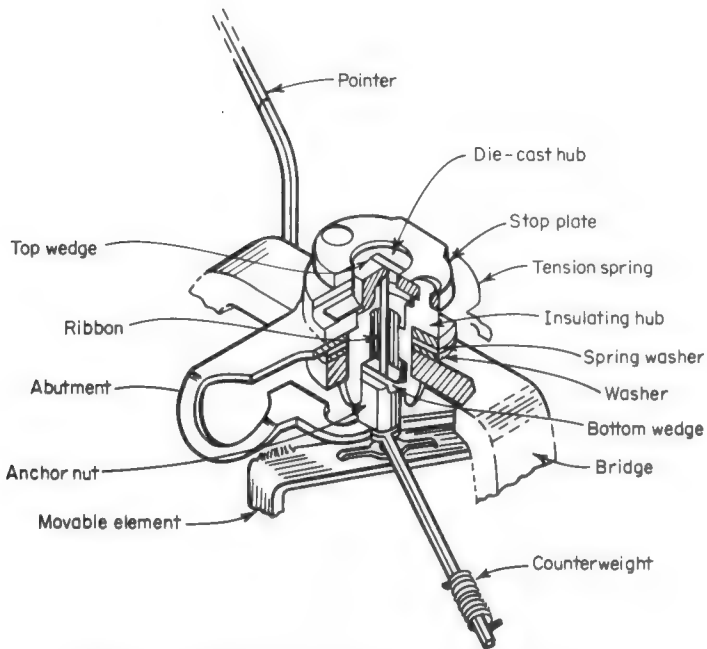
The name D'Arsonval meter movement is often used to describe the moving-coil meter movement of Fig. 7.11. In the strictest sense of the word, this is not a D'Arsonval type of movement. However, the terminology appears now to be firmly established. This type of meter movement is also referred to as a *jeweled-type* instrument because the pivot of the coil rests in tiny near-frictionless jewels. Instead of a jeweled pivot, many of the newer meters use what is known as a *taut-band movement*. Physically, the taut band is a flat metal band that twists as the pointer moves upscale.

Figure 7.12 shows the construction of a taut-band meter. The force that returns the

**7-10 Meters and Measurements**



**Fig. 7.11** A moving-coil meter and the parts that combine to make it.



**Fig. 7.12** Construction details of a taut-band moving-coil meter.

pointer to zero is due to the twist of the taut band. Taut-band meters are usually more accurate than jeweled types because of the reduction in the number of moving parts.

**METERS THAT DEPEND ON THE HEATING EFFECT OF A CURRENT** The meter movements that have been discussed so far depend for their operation on the interaction of magnetic fields. In general, the magnetic field surrounding a wire or a coil carrying an unknown current reacts with a magnetic material or a magnetic field in such a way that the attraction or repulsion moves a meter pointer. We will now discuss a type of meter that depends for its operation on the heat produced by a current flow.

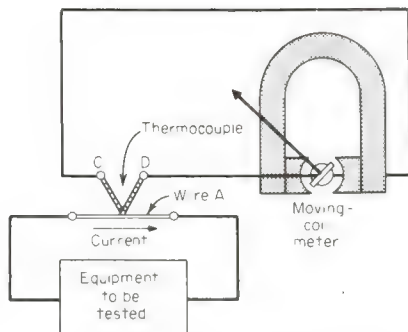


Fig. 7.13 A thermocouple ammeter.

The instrument of Fig. 7.13 is called a *thermocouple meter*. (A thermocouple is simply a junction of two dissimilar metals. When the junction is heated, a voltage is generated across the junction.) In the meter movement of Fig. 7.13, the junction is heated by current flowing in wire A, and the voltage appears across points C and D. The amount of voltage is proportional to the amount of heat at the junction of the thermocouple. This voltage generated by the thermocouple causes a current to flow through a moving-coil meter. A moving-coil meter is used here because it is one of the most sensitive available.

Wire A in the thermocouple meter movement will be heated regardless of whether a direct or an alternating current flows through it. An especially valuable application of this instrument is for measuring r-f current. R-f currents flowing through the wire cause it to heat, and therefore the current can be measured by the action of the thermocouple. However, since the r-f current flows through a straight piece of wire, there are no reactive effects that would change the magnitude of the current. A simple moving-coil meter cannot be used to measure radio frequency because the inductance in the moving coil would cause an inductive reactance in the circuit that would change the amount of r-f current flow.

## 7.6 METER CIRCUITS

The meter movements described in the previous section are designed for measuring very small currents. In order to measure larger currents, or to measure voltage and resistance, the meter movement must be placed in a circuit that will protect it from an *overload*—that is, from an excessive current which would destroy the meter movement. The most frequently encountered meter movement is the moving-coil permanent-magnet type. A typical movement of this type will deflect to its maximum full-scale reading when there are only  $50\ \mu\text{A}$  flowing through it. Obviously, it cannot be used directly for most of the measurements required for servicing electronic equipment.

**CALCULATION OF METER SHUNTS** Figure 7.14 shows the circuit for converting a sensitive microammeter to read larger current values, such as amperes and milliamperes. The meter  $M$  and its internal resistance  $R_M$  are paralleled with a shunt resistor  $R_{sh}$ . The purpose of the shunt resistor is to bypass excessive current around the



meter movement. The arrows on the illustration show how the current divides. In the operation of the circuit, the portion of the current that flows through the meter does not exceed the value required for full-scale deflection. If the value of the shunt resistance was equal to the value of the meter resistance, then the current would divide evenly so that half the current being measured would flow through each branch. However, for measuring large currents, such as an ampere, it is necessary to divide the currents in such a way that the amount flowing through the shunt resistance is, by far, *greater* than the amount flowing through the meter. Thus, the shunt resistor has a very low resistance value—often as low as a fraction of an ohm. At first appearance, the shunt may often seem to be a short circuit across the meter movements.

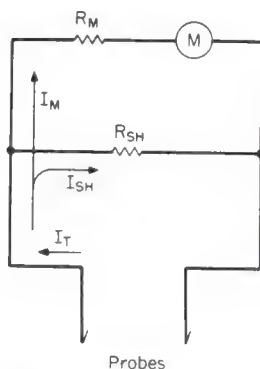


Fig. 7.14 A circuit for reading milliamperes or amperes.

The resistance value of  $R_{sh}$  (Fig. 7.14) can be calculated by application of Ohm's law and basic reasoning. The current being measured,  $I_T$ , will divide so that part of it flows through the meter,  $I_M$ , and the rest flows through the shunt  $I_{sh}$ . Mathematically, according to Kirchhoff's current law, it can be stated that

$$I_T = I_M + I_{sh} \quad (7.4)$$

where  $I_M$  = maximum current that can flow through the meter, A

$I_{sh}$  = current that must flow through the shunt, A

$I_T$  = sum of  $I_M$  and  $I_{sh}$

The maximum value of  $I_M$  is fixed by the maximum reading of the meter  $M$ . Except for the small amount of current flowing through the meter, all the current being measured must flow through the shunt. By subtracting  $I_M$  from both sides of Eq. (7.4), we obtain

$$I_{sh} = I_T - I_M \quad (7.4a)$$

(The value of  $I_{sh}$  will be needed later.)

Our basic knowledge of electricity tells us that the voltage across all branches of a parallel circuit is the same. In the circuit of Fig. 7.14 the voltage across the meter branch is, by Ohm's law, the meter current times the meter resistance  $I_M R_M$ . This must be equal to the voltage across the shunt  $I_{sh} R_{sh}$ ; so we can set them equal in equation form:

$$I_{sh} R_{sh} = I_M R_M \quad (7.5)$$

where  $I_M$  and  $I_{sh}$  = the same meaning as for Eq. (7.4)

$R_{sh}$  = current that must flow through the shunt, A

$R_M$  = sum of  $I_M$  and  $I_{sh}$

Since we are interested in calculating the value of  $R_{sh}$ , we will isolate it by dividing both sides of the equation by  $I_{sh}$ :

$$R_{sh} = \frac{I_M R_M}{I_{sh}} \quad (7.5a)$$

When calculating the value of a meter shunt, we normally have the following information: the *meter resistance*  $R_M$ , the *meter current*  $I_M$ , and the *total (maximum value of) current to be measured*  $I_T$ . We usually do *not* know the value of shunt current  $I_{sh}$ ; so Eq (7.5a) cannot be used directly. However, Eq. (7.4a) tells us that  $I_T - I_M$  can be substituted for  $I_{sh}$ . Making this substitution into (7.5a) gives

$$R_{sh} = \frac{I_M R_M}{I_T - I_M} \text{ ohms} \quad (7.5b)$$

The currents must be in amperes, and the resistances in ohms. Equation (7.5b) can be used directly for calculating the required value of shunt resistance to convert a meter movement for measuring higher current values. However, instead of memorizing the equation, it is a better practice to make the calculations by the same reasoning that was used for the derivation.

**example 7.1** A certain 50- $\mu$ A meter movement has a resistance of 90  $\Omega$ . What value of shunt resistance is needed to convert the meter so that it can measure a (maximum) current of 0.5 mA?

**solution** Given:  $I_M = 50 \mu\text{A} = 0.00005 \text{ A}$ ,  $I_T = 0.5 \text{ mA} = 0.0005 \text{ A}$ , and  $R_M = 90 \Omega$ . The current through the shunt is

$$\begin{aligned} I_{sh} &= I_T - I_M \\ &= 0.0005 - 0.00005 = 0.00045 \text{ A} \end{aligned} \quad (7.4a)$$

$$\begin{aligned} R_{sh} &= \frac{I_M R_M}{I_{sh}} \\ &= \frac{0.00005 \times 90}{0.00045} = \frac{0.0045}{0.00045} = 10 \Omega \text{ Answer} \end{aligned} \quad (7.5a)$$

It is important to remember that this type of meter must be placed *in series* with the circuit current being measured. The very low value of shunt resistance—less than a fraction of an ohm in many instances—makes the meter circuit resistance negligible. Therefore, when the meter is placed in series with the current being measured, this does not noticeably affect its value.

**CALIBRATION OF CURRENT-READING METERS** Occasionally it becomes necessary to check the accuracy of an ammeter, a milliammeter, or a microammeter. The circuit of Fig. 7.15 shows how this can be done. It is based on the principle that the same current flows through all parts of a series circuit. Since meter  $M_1$  and  $M_2$  are in series, it follows that they should indicate the same amount of current flow. Meter  $M_1$  is known to be accurate, whereas  $M_2$  is being checked for accuracy. If there is a discrepancy between the two readings, then  $M_2$  can (in many cases) be adjusted to read correctly, and this process is known as calibration. Resistor  $R_2$  is a series-limiting resistor, to prevent the current through the meters from becoming excessive, as would happen if resistor  $R_1$  were accidentally adjusted to zero ohms in the circuit. Resistor  $R_1$  can be adjusted for an exact reading for meter  $M_1$ , and the reading of  $M_2$  can be compared with this value. It is not necessary to know precise values of resistors  $R_1$  and  $R_2$  since we are only interested in the current value produced by the circuit.

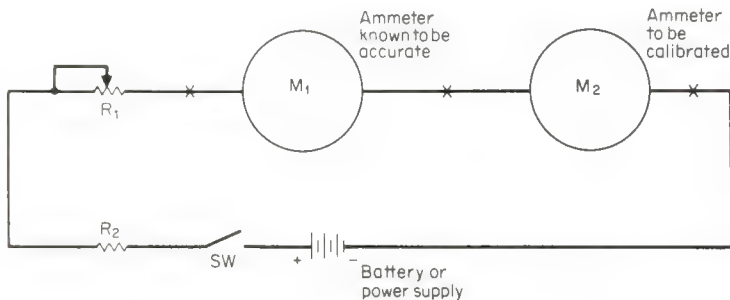
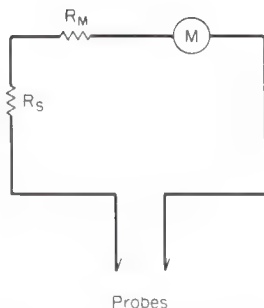


Fig. 7.15 Circuit used for calibrating ammeters.





**Fig. 7.16** A variable transformer that is useful for calibrating ac meters. (Courtesy General Radio Company)



**Fig. 7.17** A simple voltmeter circuit.

If the meters in the circuit of Fig. 7.15 are designed for measuring alternating current, then the battery should be replaced with an ac power supply. A variable power transformer, such as a *Variac* (see Fig. 7.16), is useful for this purpose. If a Variac is not available, a step-down filament transformer can be used.

**CALCULATION OF METER MULTIPLIERS** Though it is desirable for an ammeter to have an extremely low resistance so that it will not alter the circuit current value, a *voltmeter* must have a very high resistance. The voltmeter is placed across the voltage source, and ideally it will draw very little current. A typical moving-coil permanent-magnet meter movement may have a coil resistance less than  $100\ \Omega$ . If such a meter movement were placed across a 20-V source, the amount of meter current flow would be so large that it would damage the meter movement. Therefore, a series resistor  $R_S$  is placed in the circuit to limit the current flow through the meter. The series resistor is sometimes called a multiplier. Figure 7.17 shows the basic voltmeter circuit. The resistance value of  $R_S$  may be quite high, and it is often a precision resistance value.

In the circuit of Fig. 7.17, a sensitive meter movement  $M$  with an internal resistance value of  $R_M$  is placed in series with the multiplier  $R_S$ . When the probes are placed across the voltage to be measured,  $V$ , current flows in the meter circuit. This current causes a voltage drop  $V_S$  across  $R_S$  and a voltage drop  $V_M$  across the meter movement resistance  $R_M$ .

According to Kirchhoff's voltage law, the sum of the voltage drops around a circuit,  $V_M + V_S$ , must be equal to the applied voltage  $V$ . Applying this law to the circuit of Fig. 7.17 results in

$$V = V_S + V_M \quad (7.6)$$

where  $V$  = voltage being measured,  $V$

$V_S$  = resistance of multiplier,  $\Omega$

$V_M$  = resistance of meter movement,  $\Omega$

We are interested in the value of  $V_S$  at the moment. If we know the voltage across  $R_S$  (which is  $V_S$ ), and we know the current through  $R_S$  (which is the meter current  $I_M$ ), then we can calculate the resistance value of  $R_S$ .

Solving Eq. (7.6) for  $V_S$  gives

$$V_S = V - V_M \quad (7.6a)$$

Since  $V_S = I_M R_S$ , and  $V_M = I_M R_M$ , we can substitute these values into the equation to get

$$I_M R_S = V - I_M R_M \quad (7.6b)$$

where  $V$  = voltage being measured,  $V$

$I_M$  = maximum allowable meter current, A

$R_S$  = multiplier resistance,  $\Omega$

$R_M$  = resistance of meter movement,  $\Omega$

Dividing both sides of the equation by  $I_M$  to get an equation for  $R_S$ ,

$$R_S = \frac{V - I_M R_M}{I_M} \text{ ohms} \quad (7.7)$$

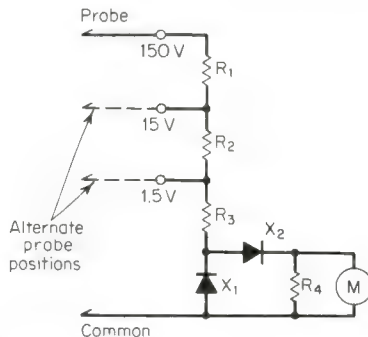
All the values on the right side of Eq. (7.7) are normally known for determining the value of a meter multiplier.

**example 7.2** A certain 50- $\mu$ A meter movement has a resistance of 90  $\Omega$ . What value of series multiplier is needed to make an instrument that will read 5 V (maximum)?

**solution** Given:  $I_M = 50 \mu\text{A} = 0.00005 \text{ A}$ ,  $R_M = 90 \Omega$ , and  $V = 5 \text{ V}$ . Therefore,

$$\begin{aligned} R_S &= \frac{V - I_M R_M}{I_M} \\ &= \frac{5 - (0.00005 \times 90)}{0.00005} = 99910 \Omega \end{aligned} \quad \text{Answer}$$

If a moving-coil permanent-magnet meter movement is used in an ac voltmeter circuit, provision must be made for converting ac to dc voltage. You will remember that the pointer of this type of meter deflects in a direction that depends upon the direction of current flow through it. The circuit of Fig. 7.18 shows an ac voltmeter arrangement



**Fig. 7.18** A meter circuit for making ac voltage measurements.

for three different ranges of voltage values. In this arrangement, the meter range is determined by the point where the probe is inserted into the circuit. For example, when the probe is plugged into the 150-V terminal, then resistors  $R_1$ ,  $R_2$ , and  $R_3$  serve the same purpose as  $R_S$  in Fig. 7.17. When the probe is plugged into the 1.5-V terminal,  $R_3$  alone serves as the series resistor.

On one half-cycle of input alternating current, diode  $X_1$  of Fig. 7.25 conducts and places an effective short circuit across the meter. On the next half-cycle diode  $X_1$  is cut off, and current flows through  $X_2$ ,  $R_4$ , and the meter movement. Resistor  $R_4$  serves as a meter shunt.

It is desirable to make the series resistance of the voltmeter as large as possible to prevent loading the circuit that is being measured. Consider, for example, the circuit of Fig. 7.17. The value of  $R_S$  cannot be so large that it prevents the meter from deflecting on full scale when the maximum voltage is being measured. The greater the current required for full-scale deflection of  $M$ , the smaller the required value of  $R_S$ , and the greater the amount of loading of the circuit by the voltmeter.

A commonly used method of rating a voltmeter according to its sensitivity is the ohms-per-volt rating. The higher the ohms-per-volt rating, the greater the sensitivity of the meter movement, and hence the lower the loading effect of the meter. The

reciprocal of the ohms-per-volt rating gives the amount of current required for full-scale deflection of the meter movement.

$$\left( \begin{array}{c} \text{Amperes of current required} \\ \text{for full-scale deflection} \end{array} \right) = \frac{1}{\text{ohms-per-volt rating}} \quad (7.8)$$

**example 7.3** A certain meter is advertised as having a sensitivity of 50 000  $\Omega/\text{V}$ . How much current is required to deflect the meter movement to full scale?

**solution**

$$\begin{aligned} \left( \begin{array}{c} \text{Amperes of current required} \\ \text{for full-scale deflection} \end{array} \right) &= \frac{1}{\text{ohms-per-volt rating}} \quad (7.8) \\ &= \frac{1}{50\,000\ \Omega/\text{V}} = 0.000\,02\ \text{A} = 20\ \mu\text{A} \quad \text{Answer} \end{aligned}$$

## 7.7 VACUUM-TUBE VOLTMETERS

The vacuum-tube voltmeter arrangement of Fig. 7.19 causes very little loading of the circuit in which a voltage is being measured. The probe is connected directly into the grid circuit of a vacuum tube which has an extremely high input resistance. The voltage of  $E_1$  is sufficient to hold the vacuum tube at cutoff when the probes are connected together. This means that the current flow through  $M$  will be zero when the probes are short-circuited together. Capacitor  $C_1$  provides an ac bypass around the grid power supply, while capacitor  $C_2$  provides a bypass around plate power supply  $E_2$ . When a positive voltage is applied to probe  $a$ , it partially cancels the negative voltage due to  $E_1$ , and the tube begins to conduct. The amount of plate current is proportional to the amount of positive voltage at  $a$ , and therefore the meter deflection is proportional to that voltage. Resistor  $R_L$  limits the plate-current flow through the meter movement.

The positive voltage applied to probe  $a$  must never exceed  $E_1$  or the tube will conduct to saturation. Under the condition of saturation, the meter deflection will no longer be proportional to the input voltage. Therefore, when changing scales of the voltmeter, it will also be necessary to change the value of the grid voltage  $E_1$ .

The circuit of Fig. 7.19 can be used to measure either a dc or an ac voltage. When an ac voltage is applied, point  $a$  becomes alternately positive and negative. On one half-cycle, probe  $a$  will be negative with respect to probe  $b$ , and this voltage will be in series with that of  $E_1$ . Under this condition the tube is driven *beyond cutoff*. On the next half-cycle, probe  $a$  is positive with respect to probe  $b$ , producing a voltage that partially counteracts grid voltage  $E_1$ . This unblocks the tube and allows current to flow through the meter movement. Although the current that flows through the meter movement is flowing only on alternate half-cycles, the inertia of the needle prevents it from dropping back to zero during the half-cycles when no plate current flows. The

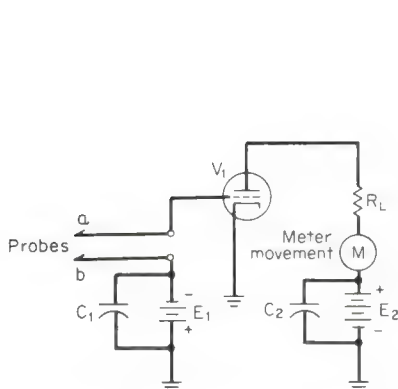


Fig. 7.19 Circuit for a simple VTVM.

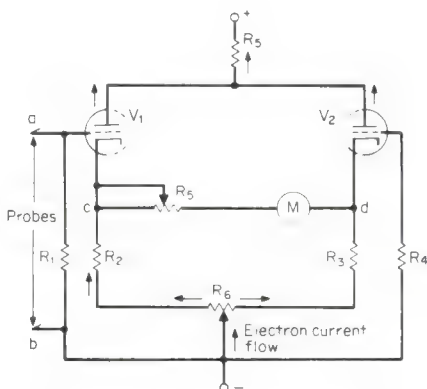


Fig. 7.20 A better VTVM circuit.

needle deflection is read on a scale that is related to the value of input ac voltage.

An alternative method of measuring ac voltage is to rectify the input voltage before it is applied to the probes. Full-wave rectification—such as obtained by a bridge rectifier—is preferable because plate current will then flow through the meter during both half-cycles of input voltage. Since the plate current does not drop to zero for a full half-cycle, it follows that, for a given ac input, the meter deflection of the meter movement will be greater for a given ac voltage than would be obtained without full-wave rectification.

In vacuum-tube voltmeters the circuit arrangement shown in Fig. 7.20 is often used. As in the case of the simple VTVM shown in Fig. 7.19, field-effect transistors may be used instead of tubes for this circuit. Tubes and FETs are interchangeable because their input impedances are quite high. Transistors, on the other hand, require current flow in their base circuit, and therefore cannot be directly interchanged with tubes or FETs.

The tube circuit of Fig. 7.20 is sometimes called a *difference amplifier*; it is also known by the names *bridge amplifier*, *differential amplifier*, and *dc amplifier*. Assuming that there is no input voltage to the probes, we see that there are two identical tube circuits,  $V_1$  and  $V_2$ . The resistance  $R_1$  and  $R_4$  are equal,  $R_2$  equals  $R_3$ , and  $V_1$  and  $V_2$  are presumed to be identically matched. If it should turn out that there are slight differences in the two tube circuits, they can be compensated for by adjusting variable resistor  $R_6$  until conduction through the two tubes is identical. Under this condition, the two cathode voltages will be identical, and therefore no current can flow through  $R_5$  and  $M$ . Assume that during measurement of a dc voltage, point  $a$  is made positive with respect to point  $b$ . This will cause  $V_1$  to conduct harder than  $V_2$ , and a cathode current through  $R_2$  will be greater than that through  $R_3$ . As a result, point  $c$  becomes more positive than point  $d$ , and current will flow through the meter. The amount of deflection of the meter needle is dependent on the amount of positive voltage applied through the probe.

If an ac voltage is applied between points  $a$  and  $b$  during the measurement,  $V_1$  will conduct harder than  $V_2$  on alternate half-cycles. Assuming that  $V_1$  is biased near the cutoff point, the negative half-cycles will quickly drive it to cutoff. Thus, for all practical purposes there will be current flow only during positive half-cycles of input ac voltages. (A better arrangement is to full-wave-rectify the alternating current before it is applied to the input probe so that  $V_1$  will conduct during both half-cycles of input ac voltage. This will increase the sensitivity of the meter.)

The circuit of Fig. 7.20 is often used in modern-day electronic systems. One very important advantage of the circuit is that it is not *tube-sensitive*. In other words, as the tubes age—assuming that the effects of their aging are identical—the circuit will still be balanced. If both tubes are placed in the same envelope, they must both be replaced when one fails, thus assuring that the tubes are matched. In the circuit of Fig. 7.19, as the tube ages, the cutoff voltage changes, and furthermore the total plate current tends to decrease with aging of the cathode. Operation of the simpler circuit, then, depends somewhat upon the age of the tubes.

Let's assume that the cathode emission of the two tubes in the difference amplifier decreases by one-third. This will still mean that point  $c$  will be identical with point  $d$  when no input voltage is applied to the probes. There will, however, be some effect on the meter reading with a voltage applied, because the amount by which  $V_1$  conducts (compared to  $V_2$  conduction) will decrease somewhat with tube aging. However, it is not nearly so critical as a single tube circuit.

Differential amplifiers—employing transistors instead of tubes—are used in many integrated-circuit applications. Transistors, like tubes, tend to change their characteristics with age. By using differential amplifiers, the problem of aging is reduced. Since both transistors are mounted on the same chip, they will age equally, and their characteristics change simultaneously.

The FET voltmeter circuit shown in Fig. 7.21 is similar to the simple tube circuit shown in Fig. 7.19. (It should be reemphasized that an FET meter may also employ a difference amplifier like the one shown in Fig. 7.27.) In the circuit of Fig. 7.21 an N-channel FET is used. When there is no input voltage to the circuit, conduction in the FET is held at approximately cutoff value. The positive voltage on the drain makes the transistor ready for conduction. The negative voltage,  $E_2$ , produces a small loop



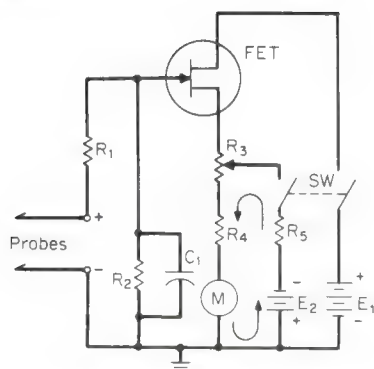


Fig. 7.21 Circuit for a FET voltmeter.

current, as shown by the arrows. This loop current exactly equals and cancels the current through the FET with no input voltage applied to the probes.

When a dc input voltage is applied to the probes in such a way that the positive terminal is at the gate, the FET will conduct. This conduction overrides the bucking current produced by  $E_2$ , and the resulting current flowing through meter  $M$  will cause the meter to deflect. The amount of deflection is proportional to the input voltage delivered to the gate. Resistor  $R_1$  is part of a voltage divider; resistor  $R_2$  forms the other part of the divider. Any input signal will appear as a voltage drop across the two resistors. The part of the voltage that appears across  $R_2$  is delivered to the gate electrode. Capacitor  $C_1$  serves as a noise signal bypass to ground.

When an ac voltage is applied to the probes, it will cause the FET to conduct on alternate half-cycles. In a better circuit arrangement, the ac voltage is full-wave-rectified with a bridge before being applied to the input probes. This would cause the FET to conduct during both half-cycles, resulting in a higher meter needle deflection for a given input ac voltage. As with other types of meters, full-wave rectification of the input ac signal being measured results in a more sensitive measuring instrument. The *sensitivity* of a measuring instrument refers to the ability of that instrument to measure very small values.

The ganged switch (SW) is used in the circuit to simultaneously remove the bucking voltage and drain voltage from the circuit when the meter is turned off, thus prolonging battery life.

An important advantage of the FET circuit over the simple circuit of Fig. 7.19 is that it does not require any filament voltage for its operation. This makes a completely portable high-impedance meter more feasible.

## 7.8 DIGITAL MULTIMETERS

All the meters discussed previously in this chapter may be referred to as "analog"-type meters. That is, they all require scales and a pointer to read a particular point on the appropriate scale. The scale readings are all continuous and change smoothly.

The analog meters all have several disadvantages in common, namely, difficulty in actual reading of the scale, the need to interpolate between printed numbers, the possibility of parallax errors, the need for multiplying factors on some ranges, limited accuracy, need for zero pointer setting, and the possibility of damage to the pointer or meter movement because of overload.

Digital multimeters overcome all the above-stated disadvantages. (A typical solid-state digital multimeter is shown in Fig. 7.2.) Such multimeters have large, easily read numbers and automatic placing of the decimal point. In some models a plus or minus sign appears at the left of the numbers, indicating the polarity of the voltage being measured. This feature is automatic and it is not necessary to switch leads or change a switch position when measuring either polarity or voltage. All measurements appear

as numbers, and obviously there is no need for interpolation, no parallax error, no need for multiplying factors, and no such thing as zero pointer setting.

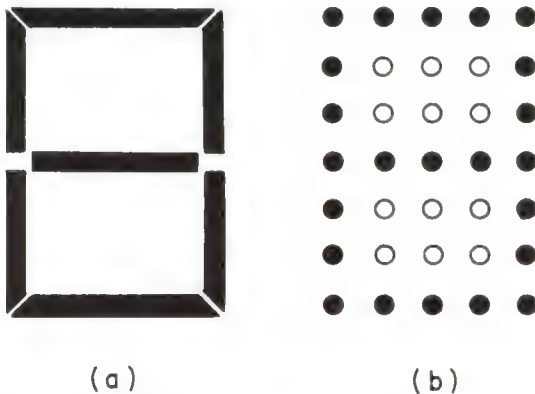
The accuracy of a digital multimeter on dc is commonly 0.1 to 1.0 percent, as compared with 2.0 percent for the usual good-quality vacuum-tube (or FET) voltmeter. Generally no damage can result to a digital meter from overload. Also, most of these meters have some sort of overload indication, ranging from a separate indicator to a scheme which causes all the digits to blink. This scheme operates in the case where the applied potential exceeds the selected scale. Some of the more expensive models also feature an "autoranger," which automatically selects the correct range required without the need to operate a range switch. Although the unit shown in Fig. 7.2 contains a read-out of four digits, other units are available with six or more digits. Obviously these latter units would be somewhat more expensive.

**DIGITAL DISPLAYS** An early type of digital display unit, which is still in fairly wide use, is the so-called "Nixie" tube. This is a small vacuum tube having a common anode and ten individual cathodes. The cathodes take the form of numbers from 0 through 9. These are stacked front to rear in the tube, which has a transparent front. The tube is a sort of glorified neon bulb. The separate cathodes are connected to the control circuitry and when a cathode is energized, its number is illuminated with an orange glow which is clearly visible. Obviously there will be as many tubes used in a particular display as the desired number of digits.

The Nixie tubes generally provide a display which is brighter and has larger numbers than the LED types (discussed below). However, they require a relatively high exciting voltage and draw too much current for practical battery operation. Thus these units are generally line-operated. The Nixie-tube display is not as visible under high ambient light conditions as the reflective liquid crystal type (discussed below).

A very popular digital display is the one which is illuminated by LEDs (light emitting diodes). This is generally found in one of two forms: the 7-segment or the 35-segment ( $5 \times 7$ ) display. These configurations are shown in Fig. 7.22. Although the number 8 is shown in both parts of this figure, any number between 0 and 9 could be shown. It is also possible to configure the displays so that a plus or minus sign may be produced. A plus sign is shown in Fig. 7.2.

The LED displays draw much less power than the Nixie types. The LEDs operate from a 1.5-V source and draw approximately 25 mA per segment. Many LED-type displays are "strobed" (intermittent, rapid lighting). This further reduces the power consumption, so that battery operation becomes very practical. The types of batteries in common use are the throwaway types, rechargeable types, and external battery packs. The readability of LED displays in high ambient light is not as good as that of the Nixie or the reflective liquid crystal types.



**Fig. 7.22** Two common methods of producing digital segmented number displays. (a) Seven-segment display showing number 8. (b) A  $5 \times 7$  (35-segment) display also showing number 8.



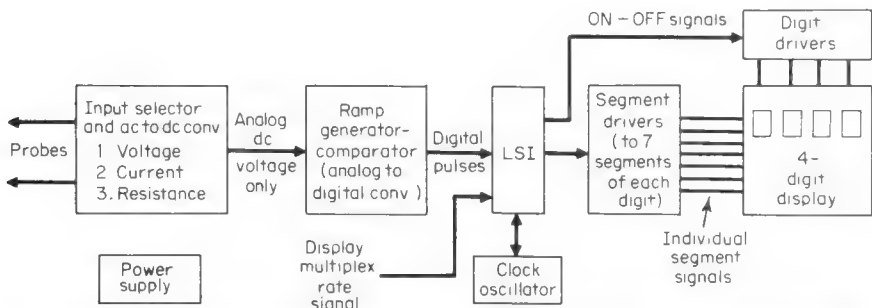
The liquid crystal displays draw by far the least current of the three types described here. Each liquid crystal segment draws only a few microwatts. Consequently this type is the most suitable for battery operation. This device is used in the seven-segment digit display shown in Fig. 7.22*a*. The liquid crystal displays come in two forms: the transmissive (back-lighted) type and the reflective (ambient-lighted) type. The transmissive type results in a greater power drain because of the back lighting and thus produces a shorter battery life. The reflective type has a very low power drain and is also used in devices such as watches and clocks, where maximum battery life is required.

Each segment of a liquid crystal display consists of two clear glass plates, with a thin layer of liquid crystal material between them. The glass plates each have conductive coatings which connect to the energizing source. When energized, the character of the liquid crystal material in the appropriate segments changes with regard to its light-transmission properties. In this manner the desired number appears, either from back lighting or from reflective lighting.

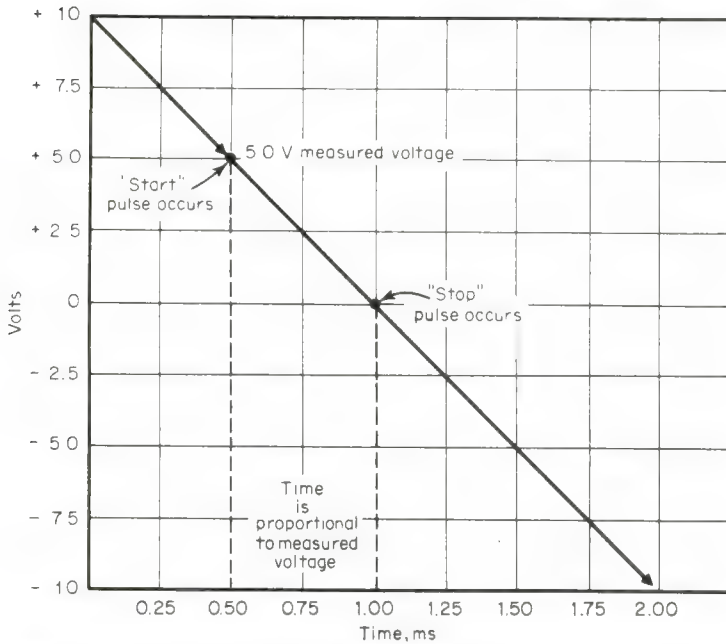
**THEORY OF OPERATION** The operation of a typical digital multimeter is explained here only in general terms. The reader should refer to Chap. 14, "Digital Circuit Fundamentals," and Chap. 15, "Digital Integrated Circuits," for the various details involved in such circuits.

The operation of a typical multimeter will be described with the aid of the simplified block diagram of Fig. 7.23. The quantity to be measured, ac or dc voltage, ac or dc current, or resistance, is fed into the input selector. The output of this selector is always a dc voltage, which is proportional to the quantity being measured. (The input selector also contains the range switching.) This proportional dc voltage is then fed to the ramp generator-comparator, which is the analog to digital converter. The output of this block will consist of digital pulses which are related to the proportional dc voltage, but only in terms of the time between them.

The ramp generator produces a linear, negative-going ramp, which begins at a positive value, goes through zero, and continues to an equally negative value. This is shown in Fig. 7.24. Note that the ramp (an upside-down sawtooth wave) voltage is proportional to time. Now assume that the voltage to be measured is 5 V. When the negative-going ramp reaches the same value as the measured voltage (5 V), the comparator puts out a pulse, often called a "start" pulse, which permits the LSI (large-scale integrated circuit) to begin counting pulses generated by the clock oscillator. The counting continues until the ramp decreases to 0 V. At this time the comparator puts out a "stop" pulse, which tells the LSI (logic circuits) to stop counting the clock oscillator pulses. Now note that the original voltage to be measured (5 V) has been translated from an analog quantity to a time interval and then to a number of clock pulses, which is exactly proportional to the original 5 V. Obviously both the ramp repetition rate and the clock oscillator frequency are quite high, to minimize the time required to display readings (about 1 ms) and to permit "strobing" (or multiplexing) of the displays. This feature will now be explained.



**Fig. 7.23** Simplified block diagram of a digital multimeter. Corresponding segments of each digit are connected in parallel.



**Fig. 7.24** When the ramp reaches the measured voltage (5 V), a “start” pulse is generated. When the ramp reaches zero volts, a “stop” pulse is generated (see text).

In the block diagram of Fig. 7.23 (and Fig. 7.2), the display is seen to consist of four digits. This means the maximum display number is 9999. However, an automatic decimal-point placement is included in the logic circuits, so that exact value is immediately evident. Note the placement of the decimal point in Fig. 7.2, where the reading is +17.76 V. When the strobing or multiplexing feature is present, as shown in Fig. 7.23, each of the four digits is lighted separately, in turn. However, this is done at such a high rate of speed that they all appear to be lighted simultaneously. Strobing saves an appreciable amount of battery power, as previously mentioned.

Now to tie it all together. The multiplexing (strobing) signal is fed to the LSI and affects both the operation of the digit drivers and the segment drivers. Thus, as one particular digit is caused to be energized, at that same exact time the segment drivers cause the proper combination of segments for that digit to appear. Since all corresponding segments of each digit are tied in parallel and connected to the segment drivers in this way, it can be seen that in order to have only one digit at a time light up, showing the correct number, the strobing or multiplexing scheme just described is necessary.

**The  $\frac{1}{2}$  digit** In Figs. 7.2 and 7.23, four-digit displays are shown, which means that the maximum number which can be displayed is 9999. However, for relatively little additional expense and circuitry it is possible to provide an additional digit in front of the above-mentioned four. This added digit can only be programmed (as in this case) to be either unlighted (no reading) or to show a 1. Thus, if energized, this so-called  $\frac{1}{2}$  digit can increase the above reading from 9999 to 19 999, or practically double the original reading. Thus a 1-V range could now read 1.9999, or a 1000-V range could now read 1 999. When the  $\frac{1}{2}$  digit is added to a four-digit display, it is called a  $4\frac{1}{2}$ -digit display. When the  $\frac{1}{2}$  digit is switched on, the meter is said to be “overranging.”

**Digital meter specifications** There are certain types of specifications which are unique to digital meters. Some of the more common ones are: sensitivity, resolution, and accuracy.

Sensitivity defines the capability of a digital voltmeter to respond to the smallest changes of measured voltage. This, of course, varies with different types. However, a rule of thumb is that the sensitivity is approximately equal to the least significant digit (extreme right-hand digit) of the reading. For example, if we are reading (as in Fig. 7.2), 17.76 V, the sensitivity would be 0.06 V.

Resolution is the inherent ability of a display to resolve one digit out of the total number capable of being displayed. For example, in a four-digit display the largest number which can be displayed is 9999 (for practical reasons, 10 000). Since the display can resolve 1 digit (least significant) out of 10 000, the resolution is 1 part in 10 000, or 0.01 percent.

The accuracy of the conventional D'Arsonal-type multimeter is expressed as a percentage of the full-scale reading. For example, a conventional meter with an accuracy rating of 2 percent, when used on the 10-V scale, would have an accuracy within plus or minus 0.2 V at full scale. However, the accuracy at lower readings is not usually known, although it must be not worse than 2 percent.

The accuracy of a digital multimeter is expressed in a somewhat more complex manner. It is given in two parts. The first part is a percentage of the actual reading; the second part (or modifier) is given as a percentage of the full scale reading, or a number of digits (least significant digit). This is best illustrated by the following three examples taken from commercial units.

*Unit A:*

DCV:  $\pm 1\%$  of reading,  $\pm 1$  digit

DCI and ACI and ACV (20 Hz–1 kHz):  $\pm 1.5\%$ ,  $\pm 1$  digit

Ohms:  $\pm 2\%$ ,  $\pm 1$  digit

*Unit B:*

DCV:  $\pm 0.1\%$  of reading,  $\pm 0.05\%$  of full scale

ACV (50–500 Hz):  $0.5\%$  of reading,  $\pm 0.1\%$  of full scale

Ohms:  $\pm 0.15\%$  of reading,  $\pm 0.1\%$  of full scale

*Unit C:*

DCV:  $\pm 0.1\%$  of reading,  $\pm 1$  digit

DCI:  $\pm 0.3\%$  of reading,  $\pm 1$  digit

ACV (45 Hz–10 kHz):  $0.5\%$  of reading, plus 2 digits

ACI:  $\pm 1\%$  of reading, plus 2 digits

## Chapter 8

# Semiconductor Devices and Transistors

### 8.1 INTRODUCTION

The variety of semiconductor devices available to the technician and engineer is prodigious. Included are rectifying, regulating, tunnel, light-emitting, light-sensing, switching, r-f, and microwave diodes. Diodes are used for converting alternating current to a unidirectional current (rectification), regulating voltage levels, detecting and mixing r-f signals, switching, clamping, in displays, and in the generation of microwave power. Some devices based partially on diode action are the silicon-controlled rectifier (SCR) and the unijunction transistor (UJT). The SCR and UJT find wide use in controlling efficiently power delivered to a load.

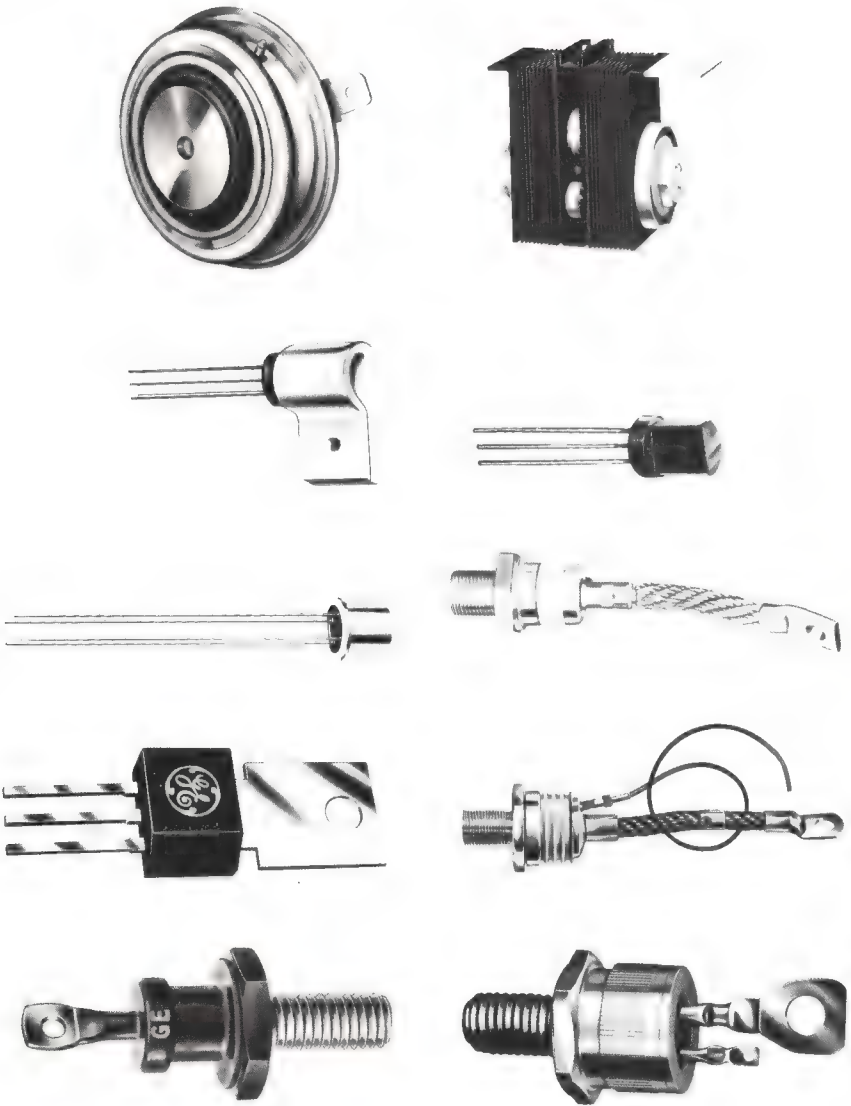
Transistors fall into two broad categories: the *bipolar junction transistor* (BJT) and the *field-effect transistor* (FET). Bipolar junction transistors, usually referred to simply as *transistors*, are available as npn and pnp types. Field-effect devices can be divided into two basic types: the *junction FET* (JFET) and the *metal-oxide semiconductor FET* (MOSFET). The MOSFET is also referred to as the *insulated-gate FET* (IGFET). Figure 8.1 provides a sampling of the available shapes and packages of discrete diodes and transistors.

In this chapter we shall examine the operation of semiconductor devices and transistors, and their characteristics, and define the significant parameters used for their characterization. A discussion of understanding diode and transistor data sheets concludes the chapter. To comprehend the operation of semiconductor devices, however, an elementary knowledge of semiconductor theory is vital.

### 8.2 ELEMENTARY SEMICONDUCTOR THEORY

Electrons appear to exhibit a dual nature, behaving as *particles* at one time and as *waves* at other times. The particle picture is useful in describing the operation of transistors and most diodes. For a few devices, such as the tunnel diode, the concept of electrons being wavelike in nature seems best suited for describing their operation.

**ATOMIC STRUCTURE** According to the particle picture, the structure of an atom is analogous to a miniature solar system. In the center of the atom is the *nucleus* which contains positive-charge particles called *protons* and, for many atoms, electrically neutral particles referred to as *neutrons*. Rotating around the nucleus are *electrons*, which have a negative charge. Because the number of electrons equals the number of protons, the atom is normally electrically neutral. The mass of the proton and neutron is approximately the same; the mass of the electron is about 1/2000th that of a proton.



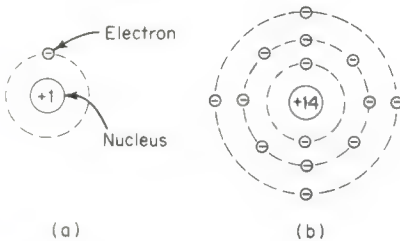
**Fig. 8.1** Discrete diodes and transistors are available in many sizes and shapes. (Courtesy General Electric Company, Semiconductor Department, Syracuse, N.Y.)

The simplest atomic structure is hydrogen (H), depicted in the two-dimensional sketch of Fig. 8.2a. Its *atomic number*  $Z$  is 1 because a single electron orbits around a nucleus containing a single proton. The orbit, or *shell*, nearest to the nucleus can hold no more than two electrons. Helium (He), for example, has an atomic number  $Z = 2$ . This denotes that exactly two electrons are rotating around the nucleus (which contains two protons), and the shell is completely filled. For this reason helium is *chemically inert* and does not react with other elements. An element whose shell is unfilled, such as hydrogen, reacts with other elements. The number of electrons in an

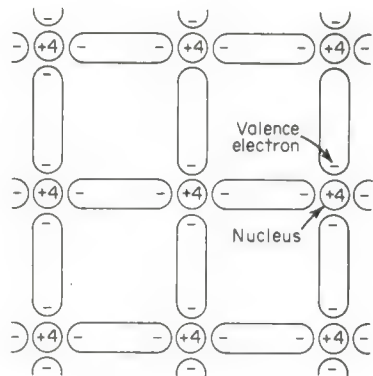


unfilled shell is called the *valence* electrons; for example, hydrogen has one valence electron.

For atoms of greater structural complexity, the electrons are contained in shells located farther away from the nucleus. The second shell can hold up to 8 electrons, the third up to 18, and so on. The atomic structure of silicon (Si), the basic material used for discrete semiconductor devices and integrated circuits, is illustrated in Fig. 8.2b. Its atomic number  $Z = 14$ ; consequently, the third shell is unfilled and contains *four* valence electrons. Germanium (Ge), another semiconductor material, also has four valence electrons. Elements containing four valence electrons are said to be *tetra-valent*.



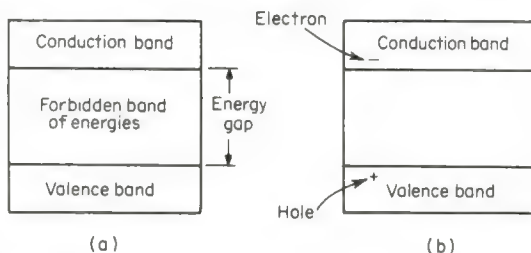
**Fig. 8.2** Examples of atomic structures: (a) Hydrogen, H. (b) Silicon, Si.



**Fig. 8.3** Representation of covalent bonding.

Silicon for semiconductor devices is grown to have a *single-crystal* structure. In simple terms, a single crystal has an ordered array of atoms. In absolutely pure, or *intrinsic*, silicon, each atom shares an electron with its four neighboring atoms, as illustrated in Fig. 8.3. This type of electron sharing results in *covalent bonds*. Because only the valence electrons are involved in covalent bonds, the silicon atom is represented simply by four valence electrons around a nucleus of +4 positive charge for electrical neutrality.

**ENERGY BANDS** The resistance of a semiconductor is greater than that of a conductor, but considerably less than that of an insulator. With the aid of an *energy-band diagram*, shown in Fig. 8.4a, one can readily understand the differences in conductors, semiconductors, and insulators. According to modern atomic theory, the energies of electrons in a material, such as silicon, are grouped in *energy bands*. The *valence band* of energies corresponds to electrons that are attracted strongly by the nucleus. As a



**Fig. 8.4** Energy-band diagrams: (a) Basic energy diagram. (b) An electron reaching the conduction band of energies leaves a hole in the valence band of energies.



result, these electrons cannot serve as carriers of current. To become carriers of current, the electrons must occupy a higher band of energies, referred to as the *conduction band*. No electron, however, can ever occupy a region, referred to as the *forbidden band of energies*, which separates the valence and conduction bands. An electron in the valence band, therefore, must have a minimum energy before it occupies the conduction band. This minimum energy is the *energy gap* of the material.

For conductors such as copper the valence and conduction bands overlap. Consequently, there is a plentiful supply of electrons to serve as carriers of current. In an insulator, such as diamond, the energy gap is so great that there are virtually no electrons in the conduction band. The energy gap for semiconductors, however, is such that some electrons are found in the conduction band.

An electron that has reached the conduction band of energies in a semiconductor leaves a vacancy, or *hole*, in the valence band of energies, as illustrated in Fig. 8.4b. What is interesting is that the hole behaves like a positively charged particle whose charge is equal and opposite to the electron. It is influenced by an electric field, like an electron, but travels in an opposite direction and more slowly than an electron.

Because of the effects of energy, such as heat or light, equal numbers of electrons and holes are produced in a semiconductor. This action is referred to as the *intrinsic generation of electron-hole pairs*. As the temperature of a semiconductor rises, the electrons in the valence band gain energy and the generation of electron-hole pairs increases rapidly. In fact, this is what limits the operating temperature of silicon devices to 150°C. Because the energy gap of germanium is nearly one-half that of silicon, the maximum operating temperature of germanium is approximately 85°C.

**DOPING** Intrinsic silicon, which contains equal numbers of electrons and holes, is useless for semiconductor devices. One needs to alter intrinsic silicon so that it has either an excess of electrons or of holes. This is accomplished by introducing suitable impurities in the silicon. The process is called *doping*, and the impurity is referred to as a *dopant*.

**n-TYPE SILICON** Assume that an impurity, such as phosphorus (P) which has five valence electrons, is introduced in silicon. (Elements with five valence electrons are called *pentavalent elements*.) The impurity concentration required for semiconductor devices is extremely minute, in the order of 1 impurity atom for  $10^8$  silicon (*host*) atoms. The impurity atom replaces a silicon atom in the single-crystal lattice. Because for every  $10^8$  silicon atoms there is only one impurity atom, the probability is high that the impurity atom will be surrounded by silicon neighbors in the lattice. Silicon doped with a pentavalent impurity is called *n-type silicon*.

A two-dimensional view of n-type silicon is illustrated in Fig. 8.5a. Because phosphorus is a pentavalent dopant, it has five valence electrons. Four of the valence electrons form covalent bonds with four neighboring silicon atoms. The fifth valence electron is free and serves as a carrier of current. When a voltage is impressed across an

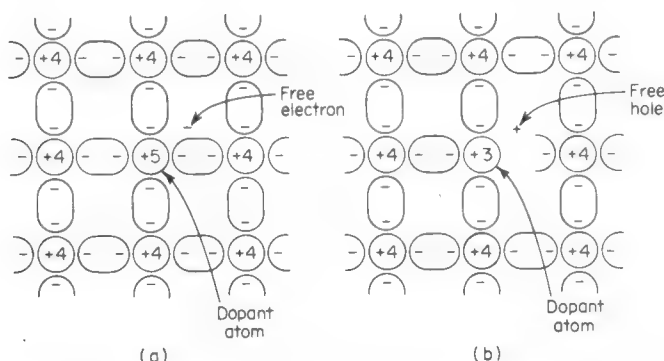


Fig. 8.5 Covalent bonding in a doped semiconductor: (a) n-type. (b) p-type.

n-type specimen, the free electrons are attracted toward the positive terminal, and current flows. If the doping is increased, more free electrons exist, and greater current flows.

Electrons in an n-type semiconductor are referred to as *majority carriers*. Regardless of the fact that the semiconductor is doped, the intrinsic generation of electron-hole pairs still continues, unabated. Hence, in addition to electrons, there are a few holes. Holes in an n-type material are referred to as *minority carriers*. At normal operating temperatures, the minority carriers are negligible. As the temperature rises, however, the minority carriers increase.

**p-TYPE SILICON** Assume that an element having three valence electrons (*trivalent element*), such as boron (B), is introduced in silicon. The result is pictured in Fig. 8.5b. The trivalent impurity now shares its three valence electrons with three neighboring silicon atoms. The deficiency of a fourth electron results in a hole. As mentioned earlier, the hole acts like a particle having a positive charge, equal and opposite to that of an electron, and serves as a carrier of current.

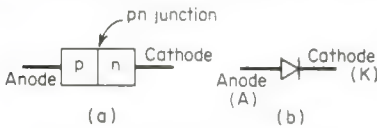
When a voltage is impressed across a p-type semiconductor, the holes are attracted to the negative terminal of the source. As for an n-type semiconductor, the greater the doping is, the more holes are present and greater current flows. Holes in a p-type semiconductor are now the majority carriers, and electrons the minority carriers.

Having this basic knowledge of semiconductor theory, we can proceed now to consider the operation of diodes and transistors.

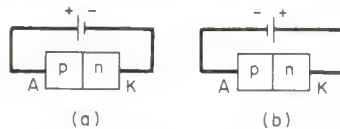
### 8.3 JUNCTION DIODE

The simplest semiconductor device is the junction diode. Its basic physical construction and electrical symbol are shown in Fig. 8.6. The junction diode is formed by doping one-half of intrinsic silicon (or germanium) with a p-type dopant and the other half with an n-type dopant. The boundary at the p and n regions, where all the action occurs, is called the *pn junction*. Connections made to the p and n regions are referred to as the *anode* and *cathode*, respectively.

When the anode is made positive with respect to the cathode, as in Fig. 8.7a, the diode is said to be *forward-biased*. Electrons in the n region are attracted toward the anode end, and holes in the p region toward the cathode end of the diode. The electrons and holes, therefore, cross the junction, and current flows in the circuit.



**Fig. 8.6** Junction diode: (a) Basic physical construction. (b) Electrical symbol.



**Fig. 8.7** Biasing a junction diode: (a) Forward-biased. (b) Reverse-biased.

In Fig. 8.7b the anode is made negative with respect to the cathode. Under this condition, the diode is *reverse*, or *back, biased*. Minority holes and electrons near the junction are initially attracted to the negative and positive terminals of the battery, respectively. As a result, the junction becomes depleted of carriers, and no further current flows. This region of depleted carriers is referred to as the *depletion region*, or *layer*. (The *voltage-variable capacitor*, or *varactor*, whose basis of operation is the depletion layer, is described in Chap. 2.)

**DIODE CHARACTERISTICS** In the previous section we saw that the junction diode behaves like a “valve.” When forward-biased, the diode permits current to flow; in the reverse-biased state, it prevents the flow of current. If it were possible to make the *perfect*, or *ideal*, diode, its characteristics might appear as shown in Fig. 8.8a.

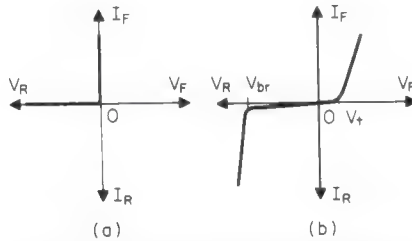
When the diode is forward-biased, current  $I_F$  (dc forward current) flows and voltage  $V_F$  (dc forward voltage) is zero, regardless of the forward current. (Note: It is common practice to use  $V$  instead of  $E$  in denoting voltage for diode and transistor characteristics.

## 8-6 Semiconductor Devices and Transistors

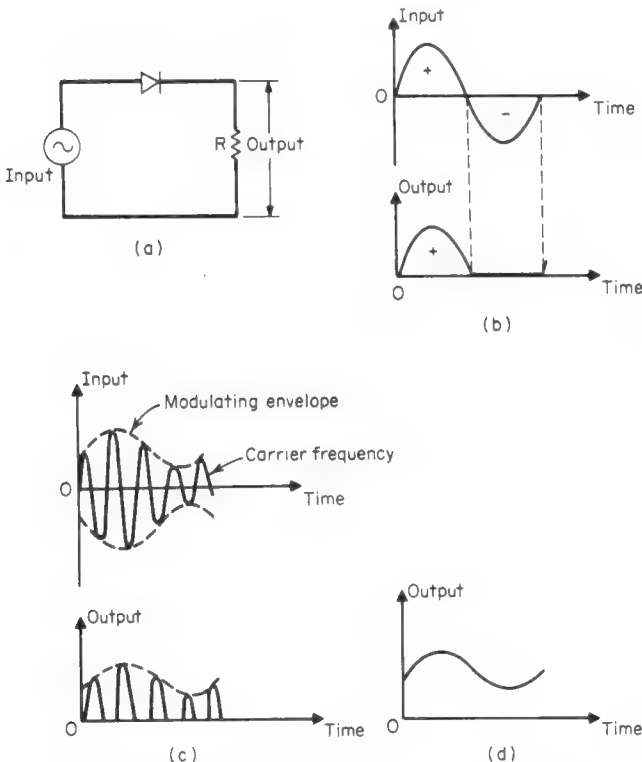
This will be adhered to in this and other chapters for describing the characteristics of semiconductor devices.) When the diode is reverse-biased,  $I_R$  (dc reverse current) is zero, regardless of the value of  $V_R$  (dc reverse voltage).

A typical characteristic curve for a *physical diode* appears in Fig. 8.8b. In forward bias, for voltages less than the threshold voltage  $V_t$ , virtually no current flows. (At room temperature,  $V_t = 0.6$  V for silicon and 0.25 V for germanium. The threshold voltage decreases at the rate of 2 mV for each degree Celsius rise in temperature.) For forward voltages greater than  $V_t$ , the diode conducts, and the voltage across the diode is somewhat greater than the threshold value.

In reverse bias, a minute reverse current, in the order of microamperes for silicon,



**Fig. 8.8** Voltampere characteristics: (a) For an ideal diode. (b) For a physical diode.



**Fig. 8.9** A diode used for rectification and demodulation: (a) Basic circuit. (b) Input and output waveforms for rectification. (c) Input and output waveforms for demodulation. (d) Modulating signal obtained from demodulator. (See Example 8.1.)

flows. The reverse current increases as the reverse bias voltage is increased. At a voltage  $V_{br}$  (the *breakdown voltage*), the reverse current increases rapidly for a very slight increase in reverse voltage. This characteristic is used in voltage-reference and regulating diodes, to be considered in the next section.

**RECTIFIER, R-F, AND SWITCHING DIODES** A diode used for rectification, or in r-f and switching circuits, is basically a pn junction diode. What distinguishes it is its physical size, geometry, and encapsulation. Because it handles many milliamperes or amperes of current, a rectifier diode is generally larger in size than a r-f or switching diode. A larger physical size implies a large junction area and capacitance. As a result, rectifier diodes cannot be used effectively in high-frequency or switching applications.

R-f diodes are physically small and designed to minimize capacitance and inductance. Switching diodes also exhibit small physical size and are designed to switch rapidly from the reverse- to the forward-biased state. Recently switching diodes have been developed that are capable of switching a few amperes in the order of a tenth of a microsecond.

**example 8.1** Show how a junction diode may be used for (a) rectification of alternating current, and (b) demodulation (detection) of an amplitude-modulated (AM) signal.

**solution** A basic circuit for both applications is illustrated in Fig. 8.9a.

(a) Referring to Fig. 8.9b for rectification, the input to the circuit is typically a 60-Hz sine wave. During the positive half-cycle, the diode is forward-biased (the anode is positive with respect to the cathode), and current flows. Neglecting the small forward voltage across the diode, the output is equal to the input. On the negative half-cycle, the diode is reverse-biased (the anode is negative with respect to the cathode), and no current flows. The output is *unidirectional*, that is, always positive. With suitable filtering (see Chap. 16), the unidirectional waveform is converted to a dc voltage.

(b) An example of an amplitude-modulated input signal is shown in Fig. 8.9c. As in rectification, on the positive half-cycle the diode conducts, and on the negative half-cycle the diode does not conduct. The output, therefore, is a group of unidirectional waves of different amplitudes. If a suitable capacitor is placed in parallel with resistor  $R$  in Fig. 8.9a, the modulating signal is obtained, as shown in Fig. 8.9d.

**DIODE PARAMETERS** Commonly used parameters for characterizing junction diodes are summarized in Table 8.1. In specifying their values, the manufacturer must

**TABLE 8.1 Commonly Used Parameters for Junction Diodes**

Parameter	Symbol	Meaning
Dc forward voltage	$V_F$	Voltage across a forward-biased diode (anode positive with respect to cathode)
Dc forward current	$I_F$	Direct current flowing in a forward-biased diode
Dc reverse voltage	$V_R$	Voltage across a reverse-biased diode (anode negative with respect to cathode)
Dc reverse current	$I_R$	Leakage current flowing in reverse-biased diode
Reverse breakdown voltage	$V_{br}, V_V, PRV, PIV$	Maximum reverse voltage across diode before it breaks down
Power dissipation	$P$	Maximum power that may be dissipated in a diode
Operating junction temperature	$T_j$	Temperature of the pn junction
Capacitance	$C$	Capacitance across diode in its forward- or reverse-biased state
Reverse recovery time	$t_{rr}$	Time required for reverse current or voltage to reach a specified value after switching diode from forward- to reverse-biased state
Forward recovery time	$t_{fr}$	Time required for forward voltage or current to reach a specified value after switching diode from its reverse- to forward-biased state
Noise figure	$NF_0$	Ratio of rms output noise power of receiver in which diode is used to that of an ideal receiver of same gain and bandwidth
Conversion loss	$L_c$	Power lost in mixer diode when converting an r-f signal to an i-f (intermediate frequency) signal
Video resistance	$R_v$	Low-level impedance of a detector diode



state the test conditions and ambient (operating) temperature at which measurements were made.

## 8.4 ZENER DIODES

The reverse characteristics of a physical-junction diode illustrated in Fig. 8.8b is redrawn in Fig. 8.10a. It was noted that as the reverse voltage increases, at a certain level,  $V_{br}$ , the diode breaks down. In the operating region of reverse current from  $I_{ZK}$  to  $I_{ZM}$ , the voltage across the diode is essentially constant.

Current  $I_{ZK}$ , the knee current, is the minimum current, and  $I_{ZM}$  is the maximum current limited by device dissipation. The voltage below the knee  $V_Z$  is the zener voltage of the device. Over the operating range of reverse current, the voltage across the diode is nearly constant and equal to the zener voltage.

Diodes made to exhibit specific zener voltages are referred to as *zener diodes*; their symbol is shown in Fig. 8.10b. Zener diodes are available having zener voltages of approximately 2 to 200 volts.

A major application for zener diodes is their use as a *voltage regulator*, shown in Fig. 8.11a. The zener diode is connected across the load, which is represented by

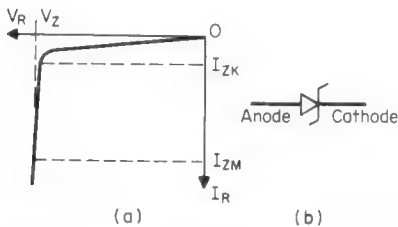


Fig. 8.10 The zener diode: (a) Voltampere characteristics. (b) Electrical symbol.

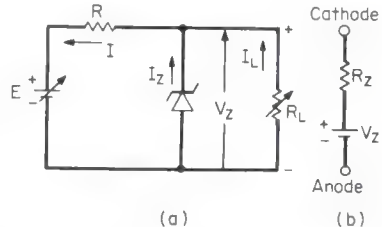


Fig. 8.11 A simple voltage regulator: (a) Circuit. (b) Model of a zener diode.

variable resistance  $R_L$ . Connected across  $R_L$  is the zener diode, which is reverse-biased (its cathode is positive with respect to the anode). The zener diode may be represented by a battery of  $V_Z$  volts in series with zener resistance  $R_Z$ , as shown in Fig. 8.11b. If the zener diode were ideal, its resistance would be zero, and the voltage across  $R_L$  would always equal the zener voltage. Because of the resistance, however, the voltage across the diode increases slightly with increasing zener current.

In Fig. 8.11a, current  $I$  flowing in resistance  $R$  always equals the sum of the zener and load currents,  $I_Z$  and  $I_L$ , respectively:  $I = I_Z + I_L$ . Assume that with  $E$  fixed,  $R$  is reduced. As a result, load current  $I_L$  increases, and the zener current decreases so that their sum is equal to  $I$ . If  $R_L$  is increased, the load current decreases and the zener current increases. During the changes in zener current, the zener voltage is hardly affected, and the voltage across the load resistance is nearly constant.

If the voltage source  $E$  should increase or decrease, the zener current will change accordingly. As before, the load voltage will hardly change its value.

**example 8.2** For the voltage regulator of Fig. 8.11a, the zener diode regulates at 20 V over a current range of 5 to 30 mA. (a) If voltage source  $E = 50$  V, determine the value of  $R$  if  $I_L$  varies from 0 to its maximum value. (b) What is the maximum load current?

**solution** (a) For zero load current, current  $I$  is equal to the maximum zener current  $I_{ZM}$ , hence,  $I = I_{ZM} = 30$  mA. The value of  $R$  equals the voltage across it ( $50 - 20 = 30$  V) divided by its current (30 mA); hence,  $R = 30 / (30 \times 10^{-3}) = 1$  k  $\Omega$ .

(b) The maximum load current equals the value of  $I$  less the minimum zener current. Therefore,  $I_{L,max} = I - I_{ZK} = 30 - 5 = 25$  mA.

**ZENER AND AVALANCHE BREAKDOWN** A reverse-biased zener diode appears to break down by two different mechanisms. For zener voltages below approximately 6 V, the high electric field across the pn junction results in the rupture of covalent bonds

and a rapid increase of electron-hole pairs. This process is referred to as zener breakdown.

In avalanche breakdown a few carriers, having acquired sufficient energy, collide with silicon atoms in the crystal lattice. As a result of the collision, new carriers are released which in turn collide with other atoms, releasing more carriers. Thus, a rapid increase of free electrons and holes occurs.

For either mechanism, the diode may be referred to as a zener or avalanche device. True zener diode action exhibits a negative temperature coefficient (breakdown voltage decreases with increasing temperature). True avalanche diode action displays a positive temperature coefficient (breakdown voltage increases with increasing temperature).

**REGULATOR AND REFERENCE DIODES** Zener diodes designed to maintain a relatively constant voltage over a wide range of load current, as in the simple regulator of Fig. 8.11a, are sometimes referred to as regulator diodes. A diode designed to exhibit a relatively fixed voltage and fixed current, regardless of temperature changes, is referred to as a reference diode.

**DOUBLE-ANODE REGULATOR** Physically, the double-anode regulator contains a single cathode which is common to two anodes (Fig. 8.12). This device is used for



**Fig. 8.12** Electrical symbol for double-anode regulating diode.

symmetrical clipping of an ac waveform and in protective circuits where negative and positive overloads may occur.

**ZENER DIODE PARAMETERS** Commonly used parameters for characterizing zener diodes are summarized in Table 8.2.

**TABLE 8.2 Commonly Used Parameters for Zener Diodes**

Parameter	Symbol	Meaning
Zener voltage	$V_z$	Nominal voltage at which zener diode regulates
Knee current	$I_{zk}$	Minimum current necessary for operation of zener diode
Maximum zener current	$I_{zm}$	Maximum current that can flow in zener diode
Zener impedance	$Z_z$	Indicates change in zener voltage for small changes in zener current with respect to a specified test current $I_{zt}$ .

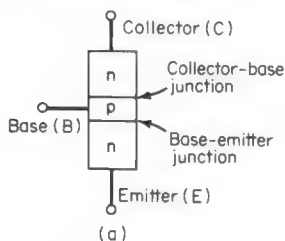
## 8.5 BIPOLAR JUNCTION TRANSISTOR

The bipolar junction transistor (BJT), usually referred to simply as a transistor, is a device with three terminals. It comes in two types: npn and pnp. In the npn device, a "sandwich" of a thin p region is surrounded by two thicker n regions as shown in Fig. 8.13a. In the pnp device, illustrated in Fig. 8.13b, a thin n region is surrounded by two thicker p regions. Electrical symbols for both types of transistors are given in Fig. 8.14.

Connected to the inner (p or n) region is the *base* (B) lead. To the outer n (or p) regions are connected the *collector* (C) and *emitter* (E) leads. For either the npn or the pnp transistor, two junctions exist: the *base-emitter* and *collector-base junctions*.

In normal operation, as for an amplifier, the BJT is connected so the base-emitter





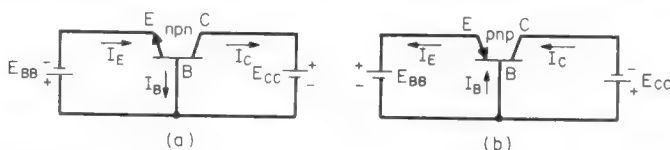
**Fig. 8.13** Basic physical construction of a bipolar junction transistor (BJT): (a) npn type. (b) pnp type.



**Fig. 8.14** Electrical symbols for a transistor: (a) npn type. (b) pnp type.

junction is *forward*-biased and the collector-base junction is *reverse*-biased. Referring to Fig. 8.15a, for an npn transistor the negative terminal of battery  $E_{BB}$  is connected to the emitter, and the positive terminal to the base. The base-emitter junction is therefore forward-biased like a pn junction diode. To the positive terminal of battery  $E_{CC}$  is connected the collector, and to the negative terminal the base. Similar to a reverse-biased *pn* junction diode, the collector-base junction is thereby reverse-biased.

To bias properly a pnp transistor, batteries  $E_{BB}$  and  $E_{CC}$  are reversed, as illustrated in Fig. 8.15b. In this case, the positive terminal of  $E_{BB}$  is connected to the emitter,



**Fig. 8.15** For amplifier operation, the base-emitter junction is forward-biased and the collector-base junction is reverse-biased: (a) For an npn transistor. (b) For a pnp transistor.

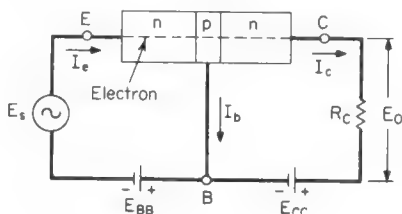
and the negative terminal to the base. The collector is connected to the negative terminal of  $E_{CC}$ , and the positive terminal is returned to the base.

For either the npn or pnp transistor, the emitter current  $I_E$  always equals the sum of the base  $I_B$  and collector  $I_C$  currents:

$$I_E = I_B + I_C \quad (8.1)$$

**OPERATION OF THE BJT** To explain the operation of a bipolar junction transistor, the npn device will be considered. The operation of the pnp transistor is identical with the npn when the word "hole" is substituted for the word "electron" in the following description.

Consider the elementary transistor amplifier of Fig. 8.16. Load resistance  $R_C$ , which, for example, can represent a loudspeaker, is connected to the collector in series with  $E_{CC}$ . The output voltage is designated by  $E_o$ . Signal source  $E_s$ , which may be the output of a phonograph cartridge, is connected in series with the base and  $E_{BB}$ . From a



**Fig. 8.16** Illustrating the operation of a common-base amplifier using an npn transistor.

signal standpoint, the batteries act like short circuits. The base, therefore, is common to the input  $E_s$  and output  $E_o$  signals. Such a configuration is called a *common-base* (CB) amplifier.

Because the base-emitter junction is forward-biased, electrons from signal  $E_s$  are readily injected into the emitter. The electrons then cross the base-emitter junction into the base region. The base, which is quite thin, has a much greater resistivity than either the emitter or the collector region. Hence, very few electrons are lost in the base, and most of them cross the base-collector junction into the collector region. The electrons then flow through resistance  $R_C$  and are attracted by the positive terminal of battery  $E_{CC}$ .

The voltage gain of an amplifier  $A_v$  is defined as the signal output voltage  $E_o$  divided by the signal input voltage  $E_s$ :

$$A_v = \frac{E_o}{E_s} \quad (8.2)$$

Output voltage  $E_o$  may be expressed as the product of output signal current  $I_c$  and load resistance  $R_C$ :

$$E_o = I_c R_C \quad (8.3)$$

Voltage  $E_s$  equals the product of input signal current  $I_e$  and the resistance of a forward-biased base-emitter junction. Letting the value of the diode resistance be designated by  $R_d$ , we have

$$E_s = I_e R_d \quad (8.4)$$

Substitution of Eqs. (8.3) and (8.4) in Eq. (8.2) yields

$$A_v = \left( \frac{I_c}{I_e} \right) \left( \frac{R_C}{R_d} \right) \quad (8.5)$$

The ratio  $I_c/I_e$  is generally designated by the Greek letter *alpha* ( $\alpha$ ) and is called the *short-circuit current gain* of a transistor in the common-base configuration. (Another symbol commonly used for  $\alpha$  is  $h_{fb}$ ). Substituting  $\alpha$  for  $I_c/I_e$ , Eq. (8.5) appears as

$$A_v = \frac{\alpha R_C}{R_d} \quad (8.6)$$

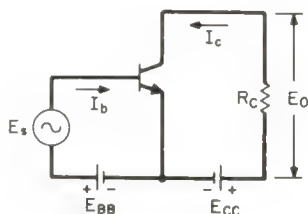
Because  $I_c$  is always slightly less than  $I_e$ ,  $\alpha$  can never exceed unity; a typical value of  $\alpha$  is 0.95. The value of  $R_C$ , however, is generally much greater than that of  $R_d$ . Therefore, even though  $\alpha$  can never exceed unity, the voltage gain can be much greater than one.

**example 8.3** In the amplifier of Fig. 8.16, assume that  $\alpha = 0.95$ ,  $R_C = 1000 \, \Omega$ , and  $R_d = 20 \, \Omega$ . Find the voltage gain.

**solution** Substituting the given values in Eq. (8.6), we obtain

$$A_v = 0.95 \times \frac{1000}{20} = 47.5$$

The most useful connection for the BJT is the *common-emitter* (CE) configuration of Fig. 8.17. In this circuit, emitter terminal  $E$  is common to the input and output



**Fig. 8.17** An elementary common-emitter amplifier using an npn transistor.

## 8-12 Semiconductor Devices and Transistors

signals. Because the base current  $I_b$  is minute, the current gain  $I_c/I_b$  is much greater than unity. The ratio, defined as the short-circuit current gain of a transistor in the common-emitter configuration, is designated by the Greek letter *beta* ( $\beta$ ) or by the commonly used symbol,  $h_{fe}$ .

**BJT PARAMETERS AND CURVES** The most commonly used parameters for characterizing the BJT are summarized in Table 8.3. Bipolar devices are classified as

**TABLE 8.3 Commonly Used Parameters for Bipolar Junction Transistors**

Parameter	Symbol	Meaning
Collector-base voltage (emitter open)	$V_{CBO}$	Maximum voltage that can be impressed across collector and base of a transistor with emitter open
Collector-emitter voltage (base short-circuited to emitter)	$V_{CES}$	Maximum voltage that can be impressed across collector and emitter of a transistor with base short-circuited to emitter. Its value is in the order of one-half of $V_{CBO}$
Collector-emitter voltage (with specified resistor between base and emitter)	$V_{CER}$	For this condition, maximum collector-emitter voltage is greater than $V_{CES}$ but less than $V_{CBO}$
Emitter-base voltage (collector open)	$V_{EBO}$	Maximum voltage that can be impressed across emitter and base of a transistor with collector open
Collector saturation voltage	$V_{CE,sat}$	Collector-emitter voltage of a transistor that is fully conducting, as in a transistor switch
Small-signal input resistance	$h_{ib}, h_{ie}$	Input resistance of a transistor with output short-circuited for the signal. This and other $h$ parameters are called <i>hybrid parameters</i> . The second subscript refers to the transistor configuration: $b$ for common base and $e$ for common emitter.
Small-signal output admittance	$h_{ob}, h_{oe}$	Output admittance of a transistor with input open-circuited for the signal
Small-signal reverse-voltage transfer ratio	$h_{rb}, h_{re}$	Ratio of voltage developed across input of a transistor to voltage present at output, with input open-circuited for the signal
Small-signal forward-current gain	$h_{fb}, h_{fe}$	Ratio of output to input signal currents with output short-circuited
Dc forward-current gain	$h_{fE}$	Ratio of dc collector current to dc base current for transistor in common-emitter configuration
Collector dissipation	$P_C$	Power dissipated in collector of a transistor. It is equal to the product of the dc collector current and dc collector-emitter voltage.
Gain-bandwidth product	$f_T$	Frequency at which common-emitter forward current gain is unity
Cutoff frequency	$f_{hfb}, f_{hfe}$	Frequency at which $h_{fb}$ (or $h_{fe}$ ) is 0.707 times its value at 1 kHz
Collector-cutoff current (emitter open)	$I_{CBO}, I_{CO}$	Reverse saturation (leakage) current flowing between collector and base with emitter open
Collector-cutoff current (base open)	$I_{CEO}$	Reverse saturation current flowing between collector and emitter with base open
Collector-base capacitance	$C_{ob}, C_{cb}$	Transistor capacitance across collector and base. This capacitance influences, to a great degree, the high-frequency performance of a transistor amplifier

small-signal, medium-power, power, r-f, and switching transistors. Several available transistors can operate at collector voltages and currents as high as 500 V and 100 A, respectively.

A useful set of curves is the common-emitter *collector characteristics* of Fig. 8.18. On the  $y$  axis is plotted the dc collector current  $i_c$ , and along the  $x$  axis the dc collector-emitter voltage,  $v_{CE}$ . A family of curves, for different values of dc base current  $I_B$ , is drawn in the figure. The reason for plotting base current is that the BJT is a *current-operated* device. As mentioned earlier, the base-emitter junction is forward-biased for normal transistor operation. Base current therefore flows and is a key quantity in establishing the operating point of a transistor.

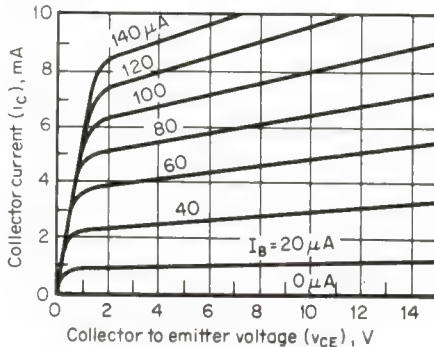


Fig. 8.18 Typical common-emitter collector characteristics.

## 8.6 FIELD-EFFECT TRANSISTOR

The field-effect transistor (FET) is basically a three-terminal semiconductor device having characteristics similar to that of a pentode vacuum tube. Unlike the bipolar transistor, the FET is a *voltage-operated* device. Instead of being biased by a current, the FET is biased by a voltage, and no input current flows. Its input resistance, therefore, is virtually infinite.

There are two kinds of field-effect devices: the *junction FET (JFET)* and the *metal-oxide semiconductor FET (MOSFET)*. The MOSFET is also referred to as an *insulated-gate FET (IGFET)* in the literature.

**JUNCTION FET** A cross-sectional view of an elementary *n-channel JFET* is shown in Fig. 8.19a. It contains two p-type regions in an n-type silicon bar. The two p regions are connected and called the *gate*,  $G$ . Ohmic connections are made to each end of the silicon bar for the *source* ( $S$ ) and *drain* ( $D$ ) leads.

A cross-sectional view of a *p-channel JFET* is shown in Fig. 8.19b. In this structure, the two n-type regions for the gate are formed in a p-type silicon bar. The biasing voltages of a p-channel JFET are opposite to those of the n-channel type (similar to npn and pnp transistors). Electrical symbols for n- and p-channel JFET's are given in Fig. 8.20.

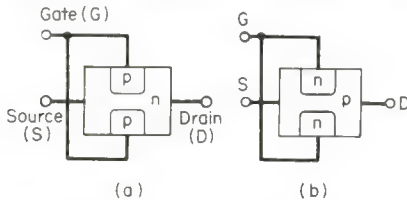


Fig. 8.19 Basic physical construction of an elementary JFET: (a) n-channel. (b) p-channel.

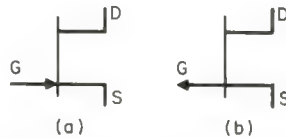
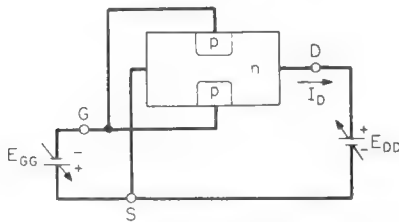
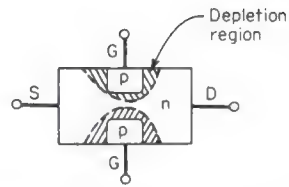


Fig. 8.20 Electrical symbols for a JFET: (a) n-channel. (b) p-channel.



**Fig. 8.21** Biasing an n-channel JFET in the common-source configuration.

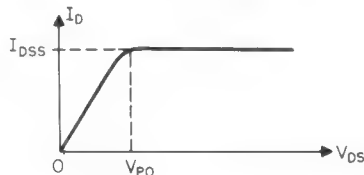


**Fig. 8.22** Depletion region developed along the channel of a JFET.

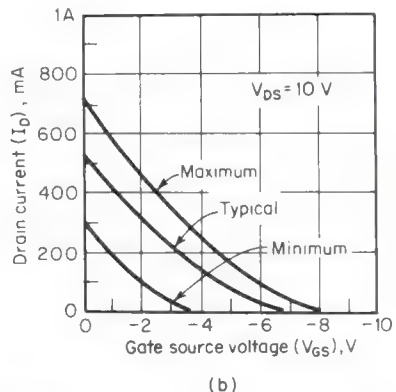
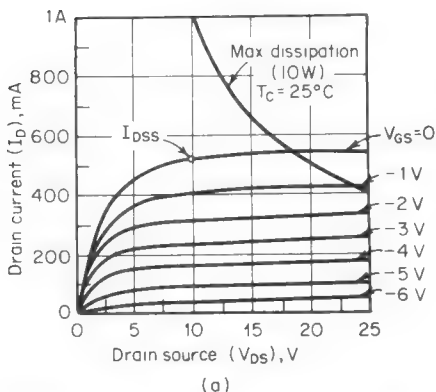
**Operation** Consider the n-channel JFET of Fig. 8.21 biased for normal operation. Because source  $S$  is common to the output (drain  $D$ ) and input (gate  $G$ ) voltages, the connection is referred to as the *common-source* configuration. The drain is held positive with respect to the source, and the gate negative with respect to the source. Assume that the gate-source voltage is initially set to zero ( $E_{GG} = 0$ ).

As the drain-source voltage  $E_{DD}$  is increased, more electrons are attracted from the source to the drain, and the drain current  $I_D$  increases. Because of the voltage drop along the  $n$  channel, a depletion region is developed along the  $pn$  junction formed by the  $p$ -type gate and  $n$ -type channel regions (Fig. 8.22). This is the same kind of depletion region that is formed in a  $pn$  junction diode. As you recall, a depletion region is devoid of current carriers. Because the channel becomes more depleted of carriers as  $E_{DD}$  is increased, the drain current levels off at a maximum value  $I_{DSS}$  (the drain current with  $E_{GG} = 0$ ), as illustrated in Fig. 8.23. Voltage  $V_{P0}$  may be defined as the minimum value of drain-source voltage,  $V_{DS}$ , for which maximum drain current flows.

As the gate-source voltage is made increasingly negative, the depletion region is



**Fig. 8.23** Drain current  $I_D$  as a function of drain-source voltage  $V_{DS}$ .



**Fig. 8.24** Typical curves for a JFET: (a) Drain characteristics. (Operation must be restricted to the area below dissipation curve.) (b) Transfer characteristics. (Courtesy Siliconix Incorporated.)



formed more rapidly than with zero bias. As a result,  $V_{PO}$  occurs at a lower value of  $V_{DS}$ . A typical family of *drain characteristics* for different values of bias voltage  $V_{GS}$  is given in Fig. 8.24a. Plotted on the  $y$  and  $x$  axes are the drain current and voltage, respectively. Because the FET is a voltage-operated device, curves for different values of gate-source bias,  $V_{GS}$ , are plotted in the figure.

Another useful curve for field-effect transistors is the *transfer characteristics* of Fig. 8.24b. This is a plot of drain current  $I_D$  as a function of gate-source bias voltage  $V_{GS}$  for a given drain-source voltage,  $V_{DS}$ .

**MOSFET** In the MOSFET, the gate is a metallic electrode separated from the channel by a thin insulating material, such as silicon dioxide ( $\text{SiO}_2$ ). There are two types of MOSFET's: *depletion* and *enhancement*. A cross-sectional view of an n-channel depletion-type MOSFET is shown in Fig. 8.25a. The silicon bar, referred to as the substrate, is p-type material. To one of the  $n+$  regions (the  $+$  sign denotes a heavily doped region) is connected the source lead, and to the other the drain lead. Between the  $n+$  regions is an  $n$  region which forms the channel.

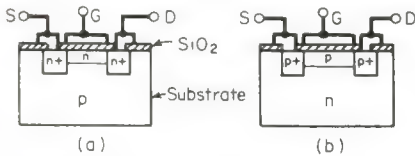


Fig. 8.25 Cross-sectional views of a depletion-type MOSFET: (a) n-channel. (b) p-channel.

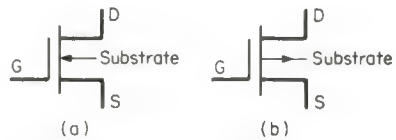


Fig. 8.26 Electrical symbols for a depletion-type MOSFET: (a) n-channel. (b) p-channel.

A cross-sectional view of a p-channel MOSFET is shown in Fig. 8.25b. The substrate in this case is n-type material, and the drain and source connections are made to p+-type material. The channel is p-type. Electrical symbols for both types of depletion MOSFET's are shown in Fig. 8.26. The substrate (SUB) is generally connected to the source terminal.

**Operation of depletion-type MOSFET** The gate, insulator, and channel of a MOSFET act like a capacitor. When, for example, a positive charge is on the gate, an opposite (negative) and equal charge is induced in the channel. If the gate has a negative charge, an equal positive charge is induced in the channel.

Consider the n-channel depletion-type MOSFET in Fig. 8.27a biased for normal operation. (A p-channel depletion-type MOSFET is biased with opposite polarities.) The drain is positive with respect to the source. If the gate is biased negatively with respect to S, a positive charge is induced in the  $n$  channel, as illustrated in Fig. 8.27b. This results in a reduction of free electrons in the  $n$  channel, and drain current  $I_D$  is reduced. As the gate is made more negative with respect to the source, the drain current is reduced further. This is illustrated in the typical drain and transfer characteristics of Fig. 8.28.

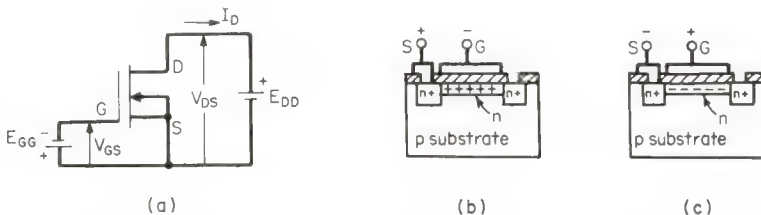
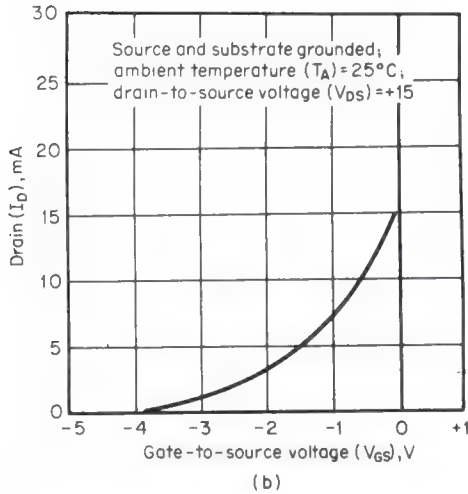
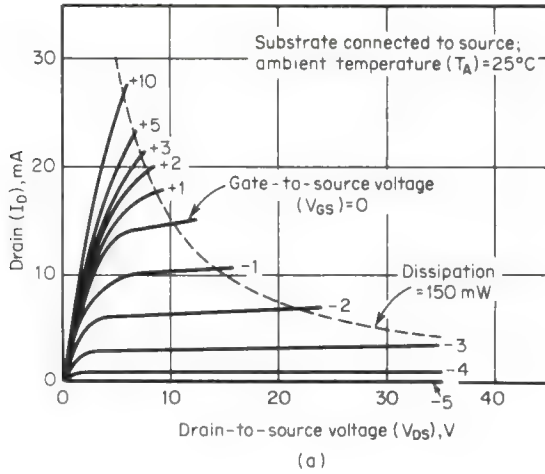
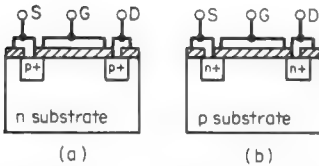


Fig. 8.27 Operation of an n-channel depletion-type MOSFET: (a) Circuit. (b) Gate biased negatively with respect to source. (c) Gate biased positively with respect to source.

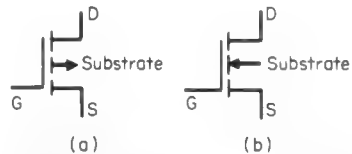




**Fig. 8.28** Typical curves for an n-channel depletion-type MOSFET: (a) Drain characteristics. (b) Transfer characteristics. (Courtesy RCA.)



**Fig. 8.29** Cross-sectional views of an enhancement-type MOSFET: (a) p-channel. (b) n-channel.

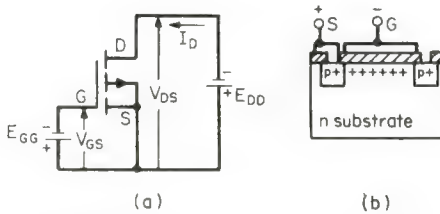


**Fig. 8.30** Electrical symbols for an enhancement-type MOSFET: (a) p-channel. (b) n-channel.

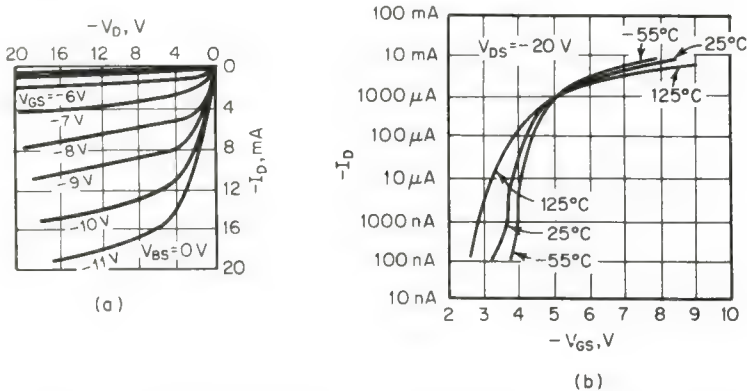
If the gate is now made positive with respect to the source, as in Fig. 8.27c, electrons are induced in the n channel. Conduction is *enhanced*, and greater drain current flows (see Fig. 8.28). Because of the insulated gate electrode, the input resistance of a MOSFET, whether the gate is positive or negative with respect to the source, approaches infinity.

**Operation of enhancement-type MOSFET** The depletion-type MOSFET may be regarded as a *normally ON* device. The reason for this designation is that for  $V_{GS} = 0$ , appreciable drain current flows (see Fig. 8.28). For many applications a *normally OFF* MOSFET is desirable, so that, for  $V_{GS} = 0$ , no drain current flows. This kind of device is realized in the *enhancement-type* MOSFET.

Cross-sectional views of p- and n-channel enhancement-type MOSFET's are illustrated in Fig. 8.29; their electrical symbols are given in Fig. 8.30. In either type, no physical channel, as in the depletion MOSFET, exists. The device, therefore, is



**Fig. 8.31** Operation of a p-channel enhancement-type MOSFET: (a) Circuit. (b) Gate biased negatively with respect to the source.



**Fig. 8.32** Typical curves of a p-channel enhancement-type MOSFET: (a) Drain characteristics. (b) Transfer characteristics. (Courtesy General Instrument Corporation.)

normally OFF. Before a current can flow between the drain and source, a suitable bias is required across the gate and source terminals.

A p-channel enhancement MOSFET biased for normal operation is illustrated in Fig. 8.31. (For an n-channel device, the polarities of sources  $E_{DD}$  and  $E_{GG}$  are reversed.) The gate is held negative with respect to the source. Positive charges (holes) are induced in the region below the gate electrode, and drain current flows. As  $V_{GS}$  is increased, the drain current also increases. Typical drain and transfer characteristics for the enhancement MOSFET are illustrated in Fig. 8.32.

**FET PARAMETERS** Commonly used parameters for characterizing FET devices are summarized in Table 8.4.

TABLE 8.4     Commonly Used Parameters for the JFET and MOSFET

Parameter	Symbol	Meaning
Drain current for zero bias	$I_{DSS}$	Drain current flowing when gate is short-circuited to source ( $V_{GS} = 0$ )
Gate reverse current	$I_{GSS}$	Leakage current flowing between gate and source for a specified reverse bias across gate and source terminals
Drain cutoff current	$I_{D,off}$	Drain current flowing when device is biased in its OFF state
Gate-source breakdown voltage	$BV_{GSS}$	Maximum reverse voltage that may be impressed across gate and source terminals without damaging the device
Gate-source pinchoff voltage	$V_p$	Gate-to-source voltage which reduces $I_{DSS}$ to 1% or less of maximum value at a specified drain-to-source voltage
Small-signal forward transconductance	$g_{fs}, g_m$	Ratio of a small change in signal drain current to a small change in signal gate-to-source voltage in common-source configuration. Parameter $g_{fs}$ is an indication of the gain for the device.
Dc drain-source ON resistance	$r_{DS}$	Ratio of dc drain-source voltage to the dc drain current, generally measured at $V_{GS} = 0$
Input capacitance	$C_{iss}$	Small-signal input capacitance for device in the common-source configuration with $V_{DS} = 0$
Reverse transfer capacitance	$C_{rss}$	Capacitance between drain and gate in common-source configuration with $V_{DS} = 0$

8.7     SILICON-CONTROLLED RECTIFIER

The silicon-controlled rectifier (SCR), also referred to as a *thyristor*, is a four-layer pnpn “sandwich” device having three terminals, as illustrated in Fig. 8.33a. Connected to the outer p region is the *anode*, and to the outer n region the *cathode*. The *gate* terminal is connected to the inner p region. The electrical symbol for the SCR is given in Fig. 8.33b.

What makes the SCR unique and useful is its switching characteristics. With the anode held positive with respect to the cathode, a small pulse of current (*trigger*) is applied to the gate terminal. The SCR turns on (*conducts*). In its conducting state, it behaves like a rectifier diode. Furthermore, it remains conducting even after the trigger is reduced to zero.

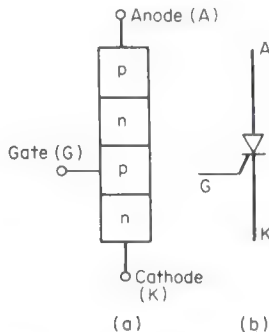
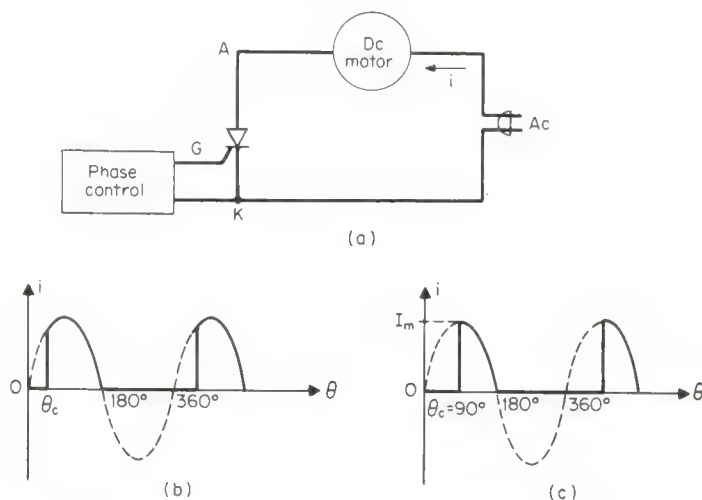


Fig. 8.33 Silicon-controlled rectifier (SCR): (a) pnpn “sandwich” construction. (b) Electrical symbol.

The SCR turns off (becomes *nonconducting*) only when the current flowing between the anode and the cathode is reduced to a level that is less than a minute critical value of current, called the *holding current*. Because of these characteristics, the SCR is extremely efficient in controlling, for example, the power to a motor or in the dimming of theater lights.

**example 8.4** Show how a SCR may be used to rectify alternating current and control the power delivered to a dc motor.

**solution** A simplified circuit for accomplishing this is illustrated in Fig. 8.34a, and the waveforms are shown in Fig. 8.34b and c. The point at which the trigger is applied during the positive half-cycle of the ac source is controlled by suitable circuitry (to be described later) in the box labeled *phase control*. Once the SCR turns on, it conducts like a rectifier diode. It stops conducting when the anode current falls below the holding current, which is close to zero. During the negative half-cycle, the SCR is reverse-biased, and no current flows.

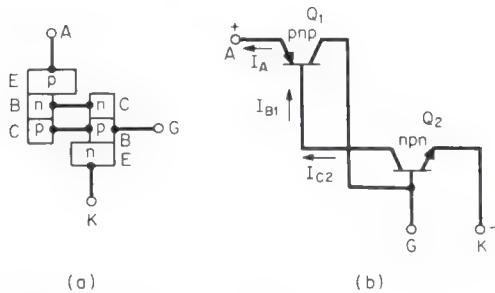


**Fig. 8.34** A SCR controlling power delivered to a dc motor: (a) Circuit. (b) Typical current waveform. (c) Current waveform for a phase angle  $= 90^\circ$ . (See Example 8.4.)

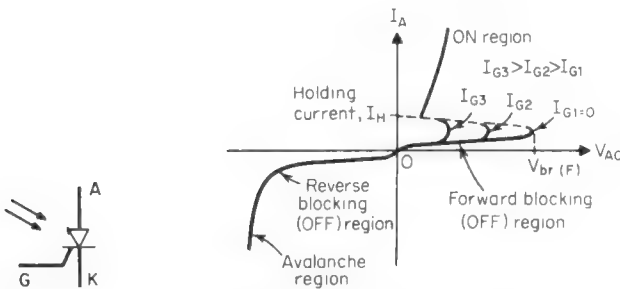
The angle at which conduction occurs, *phase angle*  $\theta_c$  (Greek letter *theta*), may be anywhere between approximately  $0$  and  $180^\circ$  (Fig. 8.34b). In the example of Fig. 8.34c, where  $\theta_c = 90^\circ$ , it is extremely simple to calculate the average (dc) current flowing in the motor. In a half-wave rectifier, current flows for  $180^\circ$  and the average current is  $0.318I_m$  amperes, where  $I_m$  is the peak current. If current flows only during half of a positive cycle ( $90^\circ$ ), the average current is therefore equal to  $0.318I_m/2 = 0.159I_m$  A.

**OPERATION OF THE SCR** For an understanding of SCR operation, it is convenient to visualize the pnpn sandwich structure of Fig. 8.33a as shown in Fig. 8.35a. The SCR is depicted here as a pnp transistor connected to an npn transistor and shown schematically in Fig. 8.35b.

Assume that anode A is positive with respect to cathode K, as indicated in Fig. 8.35b. With no positive trigger, both transistors are nonconducting, and anode current  $I_A$  is equal to a minute leakage current. Upon the application of a small positive trigger to the base of  $Q_2$  (gate terminal G), the npn transistor begins to conduct. Collector current in the npn transistor is equal to the current flowing toward the base of pnp transistor  $Q_1$ . Hence,  $I_{B1} = I_{C2}$ , and  $Q_1$  begins to conduct. Both transistors assume proper bias levels to ensure that the device conducts after the termination of the trigger. As mentioned earlier, the SCR turns off when the anode current is reduced to a value less than the holding current.



**Fig. 8.35** Viewing the SCR as being comprised of an npn and a pnp transistor: (a) Connections of the transistors. (b) Schematic representation.

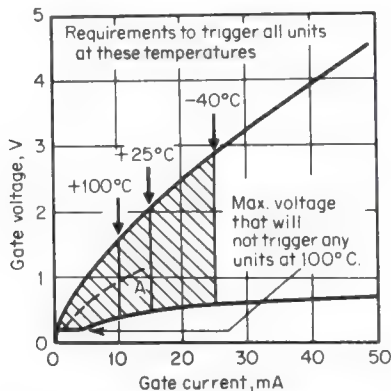


**Fig. 8.36**  
Electrical  
symbol for a  
LASCR.

**Fig. 8.37** Voltampere characteristics of a typical SCR.

**LIGHT-ACTIVATED SCR** A variation in the SCR is the light-activated SCR, referred to as the LASCR. Light passing through a translucent window in the package containing the device acts as a positive trigger. The LASCR, in other respects, behaves as a SCR. The electrical symbol for a LASCR is illustrated in Fig. 8.36.

**CHARACTERISTIC CURVES** Typical voltampere characteristics of a SCR are given in Fig. 8.37. Plotted along the vertical axis is anode current  $I_A$  and along the horizontal



**Fig. 8.38** Gate-characteristic curves for a typical SCR. (Courtesy RCA.)

axis the anode-cathode voltage  $V_{AC}$ . When  $V_{AC}$  is positive, the SCR is forward-biased. Depending on the value of gate current  $I_G$ , the device turns on, or *fires*, at a specific *forward breakdown voltage*,  $V_{(br)F}$ . Note that for  $I_{G1} = 0$ , the breakdown voltage is greater than for  $I_{G2}$  or  $I_{G3}$ , which are greater than zero. When the anode current is less than the holding current, the SCR reverts to its forward blocking (OFF) region. In the forward conducting and reverse regions of operation, the SCR resembles a junction diode.

Because the gate trigger current and voltage required vary with temperature and with units of the same type, manufacturers often supply *gate-characteristic curves* for their devices. An example of gate-characteristic curves is illustrated in Fig. 8.38. Gate voltage versus gate current is plotted for different temperatures. Note that smaller trigger voltages and currents are required at higher temperatures. For example, the maximum voltage that does not trigger the SCR at 100°C is approximately 0.2 V. Curve A represents the characteristics of a typical SCR.

**SCR PARAMETERS** Commonly used parameters for characterizing the SCR are summarized in Table 8.5.

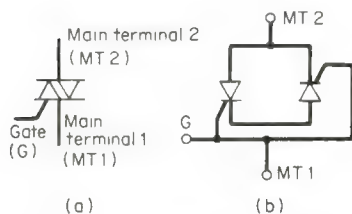
**TABLE 8.5 Commonly Used Parameters for the SCR and TRIAC**

Parameter	Symbol	Meaning
Forward breakdown voltage	$V_{(br)F}$	Forward voltage at which device fires
Reverse breakdown voltage	$V_{(br)R}$	Maximum reverse voltage that causes device to go into avalanche
On-state voltage	$V_T, V_F$	Voltage across device when it is conducting (ON state)
On-state current	$I_T, I_F$	Current flowing between anode and cathode in ON state
Holding current	$I_H$	Minimum current for device to be in ON state
Latching current	$I_L$	Minimum current to maintain device in ON state, after switching from OFF to ON state with trigger removed
Gate trigger current	$I_{GT}$	Minimum gate current for switching device from OFF to ON state
Gate trigger voltage	$V_{GT}$	Gate voltage needed to produce required gate current
Gate turn-on time	$t_{on}$	Time for device to turn on
Commutated turn-off time	$t_{off}, t_q$	Time for device to turn off

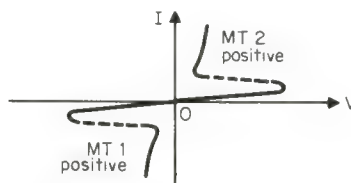
### 8.8 BIDIRECTIONAL TRIODE THYRISTOR

The electrical symbol for the bidirectional triode thyristor, or TRIAC, is shown in Fig. 8.39a, and its equivalent circuit in Fig. 8.39b. Referring to Fig. 8.39b, the TRIAC is viewed as two SCR's in parallel with the anode of one connected to the cathode of the other. The voltampere characteristics of a TRIAC (Fig. 8.40) exhibit the same forward blocking characteristics of the SCR when main terminal one is positive with respect to main terminal two, or vice versa. The parameters for characterizing the TRIAC are essentially the same as for the SCR, summarized in Table 8.5.

**example 8.5** Show how a TRIAC may be used as a lamp dimmer.



**Fig. 8.39** The TRIAC: (a) Electrical symbol. (b) Equivalent circuit.



**Fig. 8.40** Voltampere characteristics of a typical TRIAC.



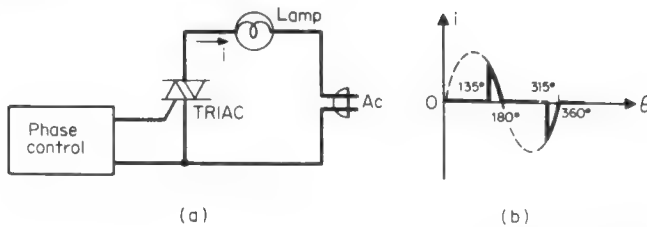


Fig. 8.41 The TRIAC used as a lamp dimmer: (a) Basic circuit. (b) Current waveform. (See Example 8.5.)

**solution** A basic circuit of a lamp dimmer is given in Fig. 8.41a. Depending on the setting of the phase control, the firing angle, as in the SCR, varies from approximately 0 to 180° in each half-cycle. For example, in Fig. 8.41b current flows only for 45° (180–135) during the positive half-cycle and another 45° (360–315) during the negative half-cycle.

## 8.9 UNIJUNCTION TRANSISTOR

The unijunction transistor (UJT) is a three-terminal device that differs from the BJT in two important respects:

1. It has only one junction.
2. It exhibits a region of *negative resistance*. A device is said to have negative resistance when, for increasing current, the voltage across the device decreases. Because of its negative resistance, the UJT finds application in oscillator, timing, and SCR trigger circuits.

The basic physical construction of a UJT is illustrated in Fig. 8.42a, and its electrical symbol is given in Fig. 8.42b. Two ohmic (nonrectifying) contacts,  $B_1$  and  $B_2$ , are made to the ends of an n-type silicon bar. The resistance of the bar, referred to as the *interbase resistance*  $R_{BB}$ , is typically of the order of 4000 to 10 000  $\Omega$ . A rectifying pn junction is formed close to terminal  $B_2$  to which is connected emitter lead  $E$ .

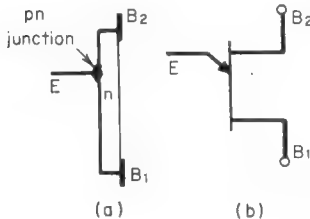


Fig. 8.42 The unijunction transistor (UJT): (a) Basic physical construction. (b) Electrical symbol.

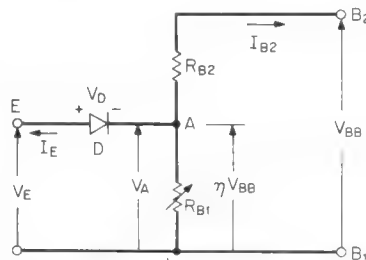


Fig. 8.43 Equivalent circuit of a UJT.

**OPERATION OF THE UJT** An equivalent circuit of a UJT is shown in Fig. 8.43. Between  $B_2$  and  $B_1$  are two resistances,  $R_{B1}$  and  $R_{B2}$ . Their sum is equal to the interbase resistance:  $R_{B1} + R_{B2} = R_{BB}$ . Because the value of  $R_{B1}$  varies as a function of emitter current  $I_E$ , it is represented as a variable resistance. The pn junction is denoted by diode  $D$  and the voltage across it by  $V_D$ . The voltage  $V_A$  at point A with respect to  $B_1$ , by voltage division, is

$$V_A = \frac{R_{B1} V_{BB}}{R_{B1} + R_{B2}} = \eta V_{BB} \quad (8.7)$$

where  $\eta$  (Greek letter *eta*) is referred to as the *intrinsic standoff ratio* of the device

$$\eta = \frac{R_{B1}}{R_{B1} + R_{B2}} = \frac{R_{B1}}{R_{BB}} \quad (8.8)$$

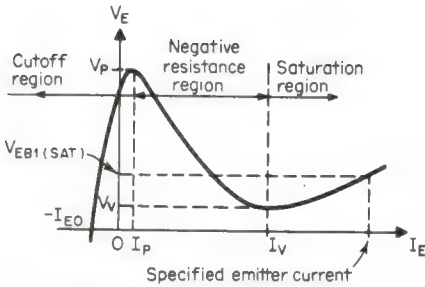


Fig. 8.44 Voltampere characteristics of a typical UJT.

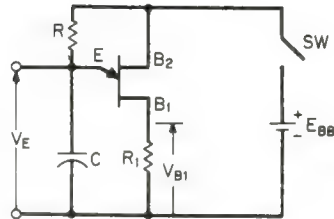


Fig. 8.45 A circuit using a UJT that generates sawtooth and trigger waveforms. (See Example 8.6.)

A plot of emitter voltage versus emitter current is illustrated in Fig. 8.44. For  $V_E = 0$ , diode  $D$  is reverse-biased, and a small reverse saturation current,  $-I_{E0}$ , flows. As  $V_E$  is increased, the diode becomes less reverse-biased and  $I_E$  less negative. At a sufficiently high emitter voltage the diode conducts and holes are injected into the bar region denoted by  $R_{B1}$ . Because of the excess holes, the value of  $R_{B1}$  decreases and  $I_E$  increases. For example, when  $I_E = 0$ ,  $R_{B1} = 4000 \Omega$ , and when  $I_E = 50 \text{ mA}$ ,  $R_{B1} = 40 \Omega$ . Owing to the reduced resistance (increased conductance) of  $R_{B1}$ ,  $V_E$  decreases as  $I_E$  increases. This behavior is referred to as *negative resistance*.

The voltage and current at the peak point of the characteristic curve are referred to as the *peak voltage*  $V_P$  and the *peak current*,  $I_P$ , respectively. Voltage  $V_P$  is the emitter voltage at which the UJT makes a transition from the cutoff region to the negative resistance region. Peak current  $I_P$  is the minimum current required to turn on the UJT.

At the valley point ( $V_V = \text{valley voltage}$ ,  $I_V = \text{valley current}$ ), the UJT makes a transition from the negative resistance region to the saturation region. In the saturation region  $R_{B1}$  acts as a positive resistance; that is, as  $I_E$  increases,  $V_E$  increases, too. Voltage  $V_{BE1, \text{sat}}$  is the voltage across the emitter and base  $B_1$  corresponding to a specified emitter current. An important application of the UJT and the significance of the peak and valley points are illustrated in Example 8.6.

**example 8.6** Analyze the operation of the UJT circuit of Fig. 8.45, and determine the waveforms across capacitor  $C$  and resistor  $R_1$ .

**solution** Assume that the capacitor is initially uncharged. When the switch is closed, the capacitor begins to charge up, through  $R$ , toward  $E_{BB}$  volts. For  $V_E$  less than  $V_P$ , the UJT is in the cutoff region. As soon as the capacitor voltage reaches  $V_P$ , the UJT is turned on and  $R_{B1}$  is reduced to a small value. The capacitor discharges rapidly through  $R_{B1}$  and  $R_1$  toward ground. But, when  $V_E = V_V$ , the UJT is turned off. The capacitor begins to charge up again and the cycle is repeated, as illustrated in Fig. 8.46a. The resulting waveform is referred to as a *sawtooth*, and its peak-to-peak amplitude is equal to the difference between the peak and valley voltages ( $V_P - V_V$ ).

The waveform across  $R_1$  is shown in Fig. 8.46b. During the time that  $C$  discharges, a current

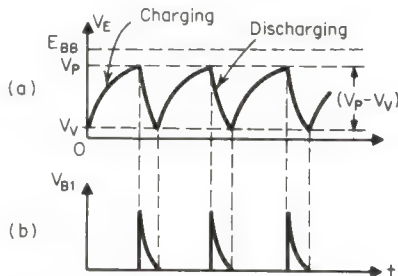
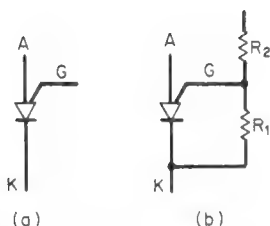
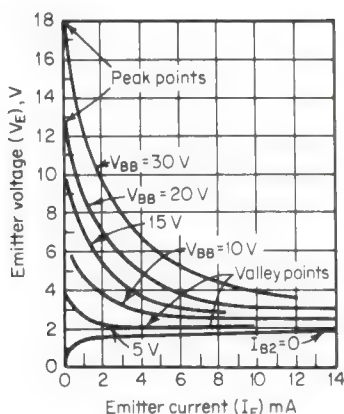


Fig. 8.46 Waveforms obtained from the UJT circuit of Fig. 8.45: (a) Across capacitor  $C$ . (b) Across resistor  $R_1$ .



**Fig. 8.47** Programmable unijunction transistor (PUT): (a) Electrical symbol. (b) Connection of resistors  $R_1$  and  $R_2$  to alter device parameters.



**Fig. 8.48** Static emitter characteristics of a typical UJT. (Courtesy General Electric Company.)

flows in  $R_1$ , and a trigger voltage is obtained which is suitable for firing a SCR. The amplitude of the trigger is determined primarily by the value of  $R_1$ . By changing the values of  $C$  and  $R$ , the time between triggers is made variable. The UJT circuit of Fig. 8.45 is commonly used as the phase control in the circuits of Figs. 8.34a and 8.41 for the SCR and TRIAC.

**PROGRAMMABLE UNIJUNCTION TRANSISTOR** The electrical symbol for the programmable unijunction transistor (PUT) is given in Fig. 8.47a. A device that is similar to the SCR, the gate of the PUT is brought out near the anode instead of near the cathode. By selecting suitable values for resistors  $R_1$  and  $R_2$  shown connected to the PUT in Fig. 8.47b, parameters such as  $\eta$ ,  $R_{BB}$ ,  $I_P$ , and  $I_V$  can be tailored to meet the user's needs. The PUT is used often in long-duration timer circuits.

**UJT CURVES AND PARAMETERS** A useful family of curves for the UJT is the *static emitter characteristics* of Fig. 8.48. Emitter voltage is plotted along the  $y$  axis and emitter current along the  $x$  axis. A family of curves for different values of interbase voltage  $V_{BB}$  is drawn in the figure. The greater the value of  $V_{BB}$ , the greater are the values of  $V_P$ ,  $I_P$ ,  $V_V$ , and  $I_V$ . Table 8.6 summarizes the important parameters for characterizing the UJT.

**TABLE 8.6** Commonly Used Parameters for the UJT

Parameter	Symbol	Meaning
Peak emitter voltage	$V_P$	Maximum emitter voltage before UJT enters negative-resistance region
Peak emitter current	$I_P$	Maximum emitter current before UJT enters negative-resistance region. It may also be thought of as minimum emitter current to turn on a UJT.
Valley emitter voltage	$V_V$	Emitter voltage at valley point
Valley emitter current	$I_V$	Emitter current at valley point
Interbase voltage	$V_{BB}$	Voltage between base one and base two
Emitter saturation voltage	$V_{BE1, \text{sat}}$	Voltage across emitter and base one at a specified emitter current and interbase voltage
Interbase resistance	$R_{BB}$	Dc resistance between base one and base two with emitter open-circuited
Intrinsic standoff ratio	$\eta$	May be defined as ratio of emitter-base-one resistance $R_{B1}$ to interbase resistance $R_{BB}$ : $\eta = R_{B1}/R_{BB}$ .

## 8.10 LIGHT-EMITTING DIODE

In a forward-biased pn junction, electrons traveling from the n region are injected into the p region. Some of the injected electrons recombine with the holes in the p region. As a result of the recombination, there is a release of radiant energy. For good efficiency and greater released energy, materials such as gallium (Ga), arsenic (As), and phosphorus (P) are used instead of silicon or germanium. A semiconductor diode designed to emit light is called a *light-emitting diode (LED)*; its electrical symbol is shown in Fig. 8.49.

The wavelength of emitted light depends on the energy gap of the material. For example, gallium arsenide phosphide (GaAsP) emits red light, and gallium phosphide (GaP) emits green light. Light-emitting diodes are available in numeral and alphabetic characters measuring in height from  $\frac{1}{8}$  to approximately  $\frac{3}{4}$  in. Their greatest application is as indicators in, for example, electronic calculators.

**CHARACTERISTIC CURVES** A number of typical characteristic curves for a LED are illustrated in Fig. 8.50. In operation, the LED is forward-biased, like the junction

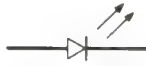


Fig. 8.49 Electrical symbol for a light-emitting diode (LED).

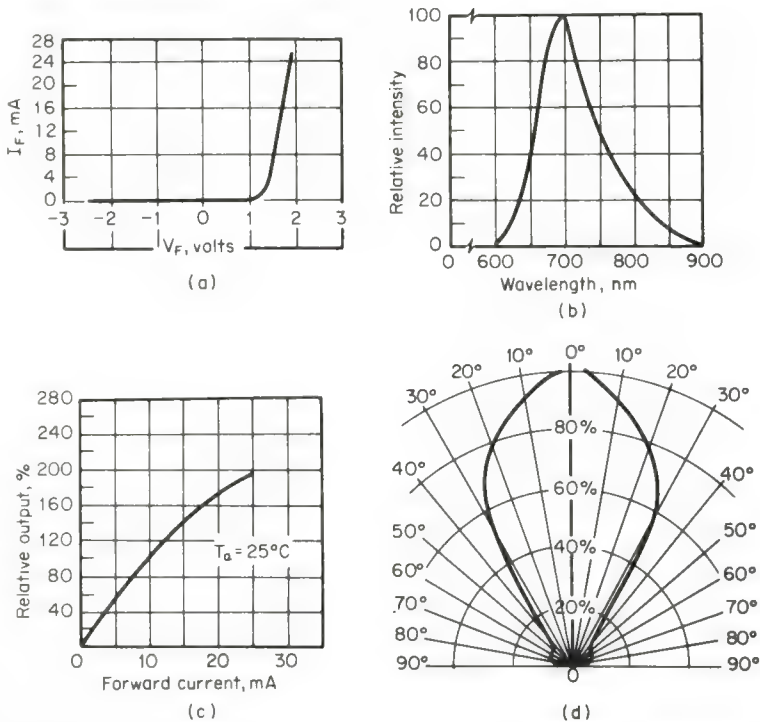


Fig. 8.50 Typical characteristic curves for a LED: (a) Voltampere characteristics. (b) Spectral distribution curve. (c) Visible output as a function of forward diode current. (d) Relative candlepower distribution curve. (Courtesy General Electric Company.)

diode. An example of the voltampere characteristics of a LED is shown in Fig. 8.50a. The threshold voltage is approximately 1.5 V.

Another curve of interest is the spectral distribution of light emanating from a LED. Such a curve is given in Fig. 8.50b. Relative intensity of the emitted light is plotted as a function of wavelength in nanometers ( $10^{-9}$  m). In this example, maximum intensity is obtained at a wavelength of 700 nm.

The dependency of the brightness of the emitted light and the forward current is illustrated by the curve of Fig. 8.50c. The curve of Fig. 8.50d shows the distribution of light for different viewing angles. *Candlepower* refers to the luminous intensity, that is, the brightness of the emitted light.

**LED PARAMETERS** Commonly used parameters for characterizing the LED are summarized in Table 8.7.

**TABLE 8.7** Commonly Used Parameters for the LED

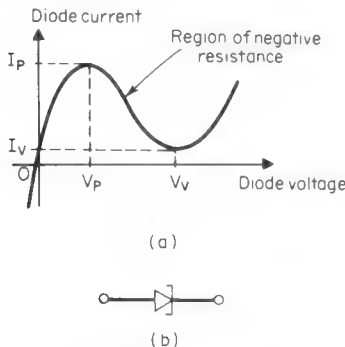
Parameter	Symbol	Meaning
Forward voltage	$V_f$	Dc forward voltage across an LED
Candlepower	CP	Measure of luminous intensity (brightness) of emitted light
Radian power output	$P_a$	Light power, or brightness, of an LED
Peak spectral emission	$\lambda_{\text{peak}}$	Wavelength of brightest emitted color. (Symbol $\lambda$ is the Greek letter <i>lambda</i> .)
Spectral bandwidth	BW $_{\lambda}$	Indication of how concentrated is the emitted color

**8.11 OTHER SEMICONDUCTOR DEVICES**

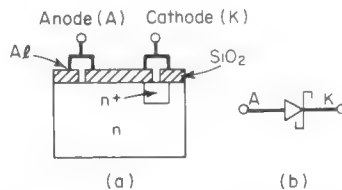
In addition to the devices described, there is a host of other semiconductor devices that are of interest. In this section, many of these devices are described and some of their important characteristics examined.

**TUNNEL DIODE** The tunnel diode, also referred to as the *Esaki diode* after its inventor, is a two-terminal device that exhibits a region of negative resistance similar to that of a UJT. Materials used in its manufacture include silicon, germanium, gallium arsenide, and gallium antimony. Typical characteristics and the electrical symbol for the tunnel diode are illustrated in Fig. 8.51. The tunnel diode may be used as an amplifier, an oscillator, or a switch, and operates at very high frequencies (in the gigahertz region).

Referring to Fig. 8.51a,  $I_p$  and  $V_p$  are the peak current and voltage, respectively;  $I_v$  and  $V_v$  are the valley current and voltage, respectively. Two important parameters for the tunnel diode are the peak-to-valley current ratio,  $I_p/I_v$ , and the frequency where the



**Fig. 8.51** Tunnel diode: (a) Typical voltampere characteristics. (b) Electrical symbol.



**Fig. 8.52** Schottky diode: (a) Cross-sectional view of its basic physical construction. (b) Electrical symbol.



negative resistance is reduced to zero,  $f_{ro}$ . For switching, the device should exhibit a high  $I_p/I_V$  ratio; for high-frequency amplifier or oscillator operation,  $f_{ro}$  should be high.

**SCHOTTKY BARRIER DIODE** The Schottky barrier diode, also referred to as a *hot carrier diode*, achieves rectification by a metal semiconductor junction. A cross-sectional view of a basic Schottky diode is illustrated in Fig. 8.52a, and its electrical symbol in Fig. 8.52b. Contacts to the substrate are aluminum, a p-type impurity. The anode A is formed by the aluminum in contact with the n-type material. A pn junction is formed, which is referred to as a Schottky barrier. The cathode is a nonrectifying contact.

The Schottky diode is an example of a *majority carrier device* because only electrons are involved in its operation. (Both electrons and holes flow in a pn junction diode.) The Schottky diodes exhibit extremely fast switching capabilities and are also useful as detectors and mixers at microwave frequencies.

Typical forward and reverse characteristics of a Schottky diode are illustrated in Fig. 8.53. Currents  $I_F$  and  $I_R$  are the forward and reverse diode currents, respectively. Voltages  $V_F$  and  $V_R$  are the forward and reverse voltages across the diode, respectively. Note that at room temperature (25°C), the threshold voltage of a Schottky diode is approximately 0.1 V; for a silicon junction diode it is 0.6 V.

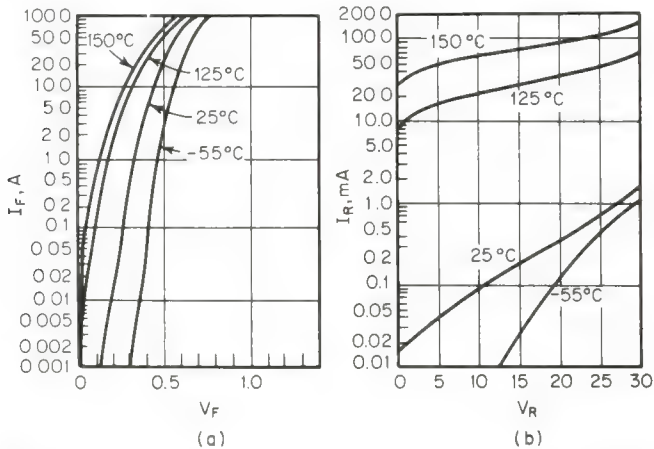


Fig. 8.53 Forward and reverse characteristics of a power Schottky diode: (a) Forward characteristics. (b) Reverse characteristics. (Courtesy TRW Semiconductor Division.)

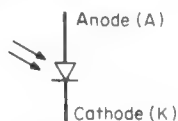
**PHOTODIODES** There are two basic kinds of photodiodes: the *photovoltaic* and the *photoconductive* types. For the photovoltaic type, no external bias source is required; an external bias, however, is needed for the photoconductive types. The electrical symbol for the photodiode is given in Fig. 8.54.

From our earlier discussion of the operation of the pn junction diode, it was seen that its reverse saturation current (electron-hole pairs) increased with rising temperature (heat). Heat is a form of energy in the infrared region of the electromagnetic spectrum. Light, in the visible region of the spectrum, striking an exposed pn junction also results in the generation of electron-hole pairs. The greater the intensity and higher the frequency of the light, the greater is the diode current.

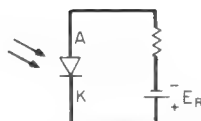
The diode is sensitive to certain colors, and its sensitivity is displayed by a spectral distribution curve similar to Fig. 8.50b. Sensitivity depends on the energy gap of the material used for the photodiode. In addition to germanium and silicon, cadmium selenide (CdSe), cadmium sulfide (CdS), and cadmium telluride (CdTe) are commonly used materials.

In connecting the conductive-type photodiode, the pn junction is reverse-biased, as





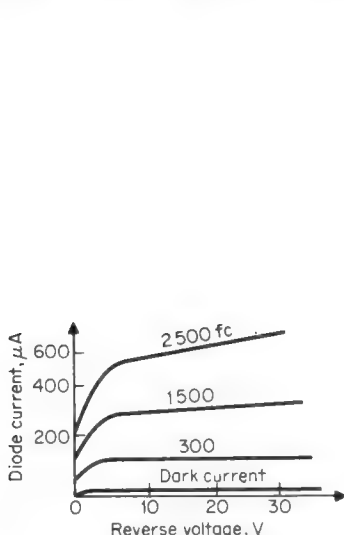
**Fig. 8.54** Electrical symbol for a photodiode.



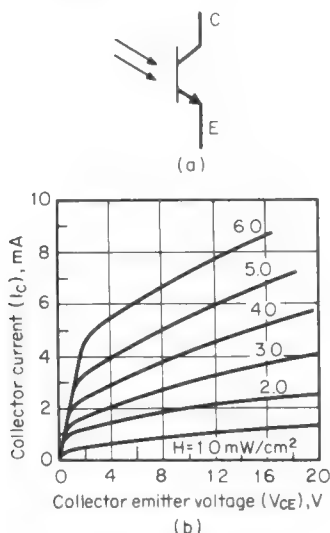
**Fig. 8.55** Biasing a conductive-type photodiode.

illustrated by the simple circuit of Fig. 8.55. Variation in diode current with reverse voltage for different levels of illumination (expressed in footcandles, fc, a unit of illumination) is illustrated in Fig. 8.56. The *dark current* is a minute leakage current that flows for no incident radiation.

**PHOTOTRANSISTOR** The phototransistor operates like a photodiode, and, at the same time, the current generated by the light is amplified. The electrical symbol of a phototransistor is shown in Fig. 8.57a. One may view the device as an npn transistor with the incident radiation replacing the base of a BJT. Typical collector characteristics of a phototransistor are given in Fig. 8.57b for different levels of illumination. In this set of curves, illumination is expressed in milliwatts per square centimeter.



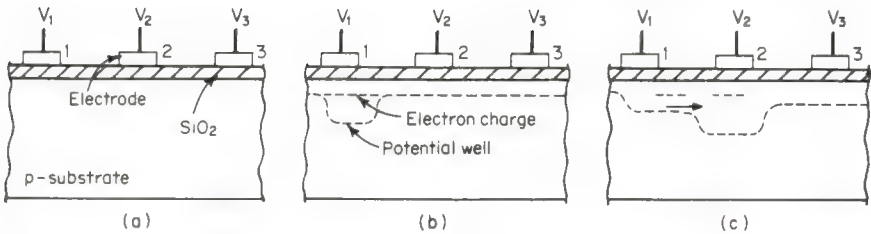
**Fig. 8.56** Voltampere characteristics of a typical conductive-type photodiode.



**Fig. 8.57** Phototransistor: (a) Electrical symbol. (b) Typical collector characteristics. (Courtesy Motorola Semiconductor Products, Inc.)

**OPTO-ISOLATOR** The opto-isolator, also referred to as an *optoelectronic coupler*, generally consists of an infrared LED and a silicon phototransistor combined in a single package. A significant advantage of this component is its high isolation resistance (in the order of  $10^{11} \Omega$ ) between its input and output. Applications for the opto-isolator include the interfacing of different types of logic circuits and their use in level- and position-sensing circuits.

**CHARGE-COUPLED DEVICE** A cross-sectional view of a charge-coupled device (CCD) is illustrated in Fig. 8.58a. Structurally, it consists of a p- (or an n-) type silicon



**Fig. 8.58** Charge-coupled device (CCD): (a) Cross-sectional view of its basic physical construction. (b) Potential well and electron charge under electrode 1. (c) Electron charge being transferred to under electrode 2.

substrate over which is a layer of silicon dioxide ( $\text{SiO}_2$ ). On top of the silicon dioxide is an array of metallized electrodes which are connected to signal voltages. In Fig. 8.58,  $V_1$ ,  $V_2$ , and  $V_3$  constitute a three-phase system of voltages.

Assume that  $V_1$  is greater than either  $V_2$  or  $V_3$ . A *potential level*, or *well*, is formed and an electron charge exists below electrode 1, as shown in Fig. 8.58b. If  $V_2$  is made greater than  $V_1$ , the charge moves and resides below electrode 2, as illustrated in Fig. 8.58c. In this manner, the charge in the substrate is transferred under one electrode to the next, and so on.

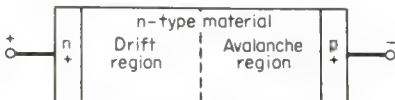
The CCD finds application as a dynamic shift register in computers and in solid-state imaging, such as video cameras. The packing density (the number of devices occupying the substrate) is in the order of 100 times as great as for other semiconductor devices. The CCD consumes very little power and is capable of operating at high frequencies.

**MICROWAVE POWER DIODES** There are two basic families of diodes used in the generation of microwave frequencies: the *avalanche-type* and the *transferred-electron bulk device*. An example of a diode that operates in the avalanche mode is the IMPATT (IMPact Avalanche and Transit Time). The LSA (Limited Space-charge Accumulation) is representative of the transferred-electron bulk device.

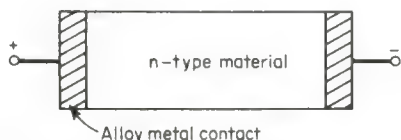
A cross-sectional view of an IMPATT diode is given in Fig. 8.59. An n region is sandwiched between a thin n+ and p+ region. Generally the device is fabricated from gallium arsenide. In operation, the diode is reverse-biased, and, at a sufficiently high dc level, the n material is divided into two regions: the drift and avalanche regions. Because of the high electric field, electron-hole pairs are generated essentially in the avalanche region. The flow of carriers is proportional to the voltage in the drift region.

At the oscillating frequency, the phase shift between the two regions is  $90^\circ$ , and the device exhibits a negative resistance. When it is connected to a tuned circuit, the negative resistance may be thought of as canceling out the positive resistance of the tuned circuit, and the combination oscillates.

A cross-sectional view of an LSA diode is illustrated in Fig. 8.60. An n-type region is sandwiched between two alloyed metal contacts; the n-type material is gallium arsenide. Above a given threshold voltage, a charge of electrons (space charge) in the material is dissipated before it can build up appreciably. The current, therefore, tends to decrease as the impressed voltage is increased. This results in the device displaying negative resistance and serves as an oscillator, as for the IMPATT diode.



**Fig. 8.59** Cross-sectional view of a basic IMPATT diode.



**Fig. 8.60** Cross-sectional view of a basic LSA diode.

## 8.12 SEMICONDUCTOR CHIPS

With the growth of hybrid technology (see Chap. 9), *unencapsulated* diodes, transistors, and even monolithic integrated circuits are available for use in hybrid microelectronic circuits. The unencapsulated devices are referred to as *chips*. There are three basic types of chips: *flip chip*, *beam lead*, and *leadless inverted devices*. Miniature encapsulated devices, which may measure less than  $0.1 \times 0.1 \times 0.05$  in, are also used occasionally.

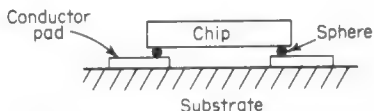


Fig. 8.61 Connecting flip chip to conductor pads on a substrate.

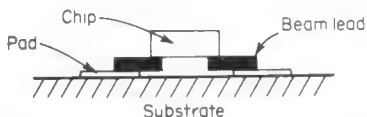


Fig. 8.62 Connecting beam-lead device to conductor pads on a substrate.

**FLIP CHIP** Contacts on the flip chip are on one side of the device and may take the form of solder-coated metal pillars or spheres about 0.05 in. in diameter. Figure 8.61 shows how a flip chip is connected to conductor pads on a substrate. By heating the pad in contact with the chip, solder flows between the pad and sphere and an electric connection is made.

**BEAM-LEAD DEVICE** In a beam-lead device, relatively long leads, or beams, are attached to the semiconductor chip (see Fig. 8.62). The beam-lead device is connected to the pads on a substrate in a manner similar to that for a flip chip.

**LEADLESS INVERTED DEVICE** The leadless inverted device (LID), also referred to as a *ceramic flip chip*, is a ceramic block in which is mounted the semiconductor chip (Fig. 8.63). Wires from the chip are bonded to the metallized lands. The LID is positioned over the pads on the substrate, and a connection is made with the application of heat.

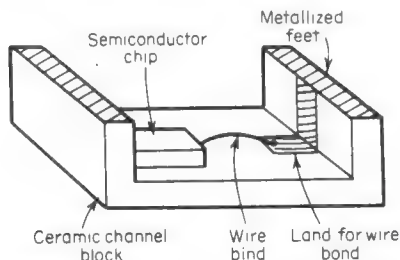


Fig. 8.63 Constructional features of an LID.

## 8.13 UNDERSTANDING DATA SHEETS

To develop an understanding of data sheets for semiconductor devices, three typical examples will be examined. The data sheets cover a medium-current silicon rectifier (Fig. 8.64), a germanium pnp transistor (Fig. 8.65), and an n-channel JFET (Fig. 8.66). The parameters specified on the sheets have been defined in appropriate tables contained in the chapter.

A study of the sheets reveals a commonality in the specifications of data and types of curves. Each data sheet has these common features:

1. A brief statement of the intended application and outstanding features of the device. For example, the rectifier is "... desirable in industrial applications that require high surge capabilities." The transistor is "... intended for industrial general-

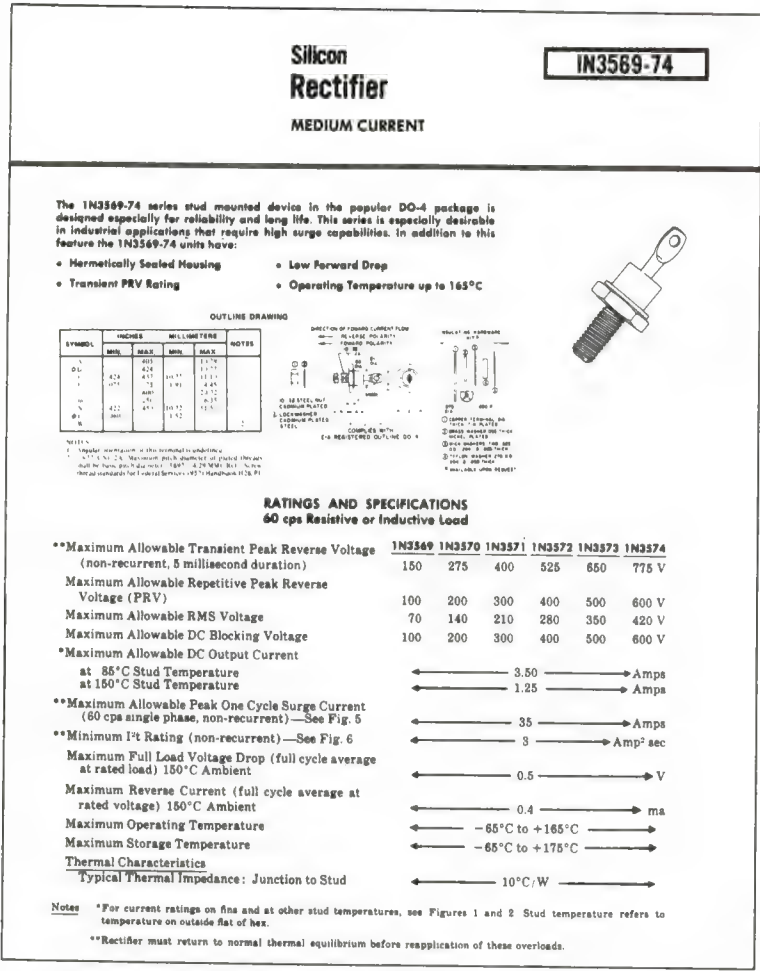
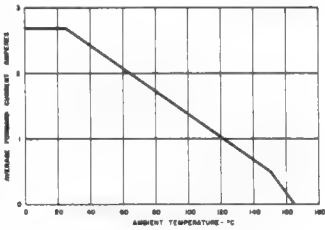
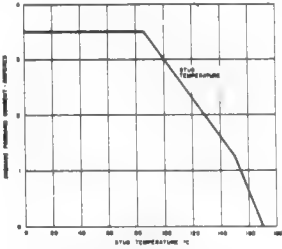


Fig. 8.64 Typical data sheet for a rectifying diode. (Courtesy General Electric Company.)

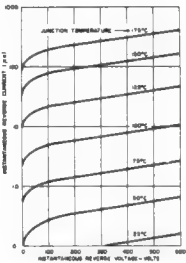
**IN3569-74**



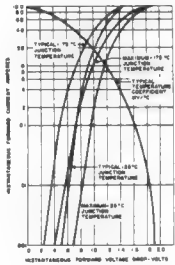
1. MAXIMUM ALLOWABLE DC OUTPUT CURRENT AS A FUNCTION OF AMBIENT TEMPERATURE FOR 1/4" SQUARE COOLING PIN-SINGLE PHASE RESISTIVE OR INDUCTIVE LOAD.



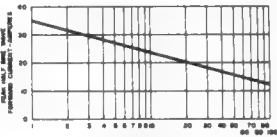
2. MAXIMUM ALLOWABLE DC OUTPUT CURRENT AS A FUNCTION OF STUD TEMPERATURE SINGLE PHASE RESISTIVE OR INDUCTIVE LOAD.



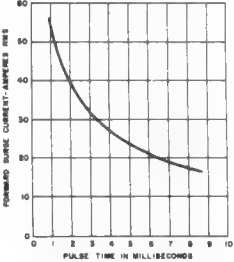
3. TYPICAL REVERSE CHARACTERISTICS



4. TYPICAL AAS MAXIMUM FORWARD CHARACTERISTICS

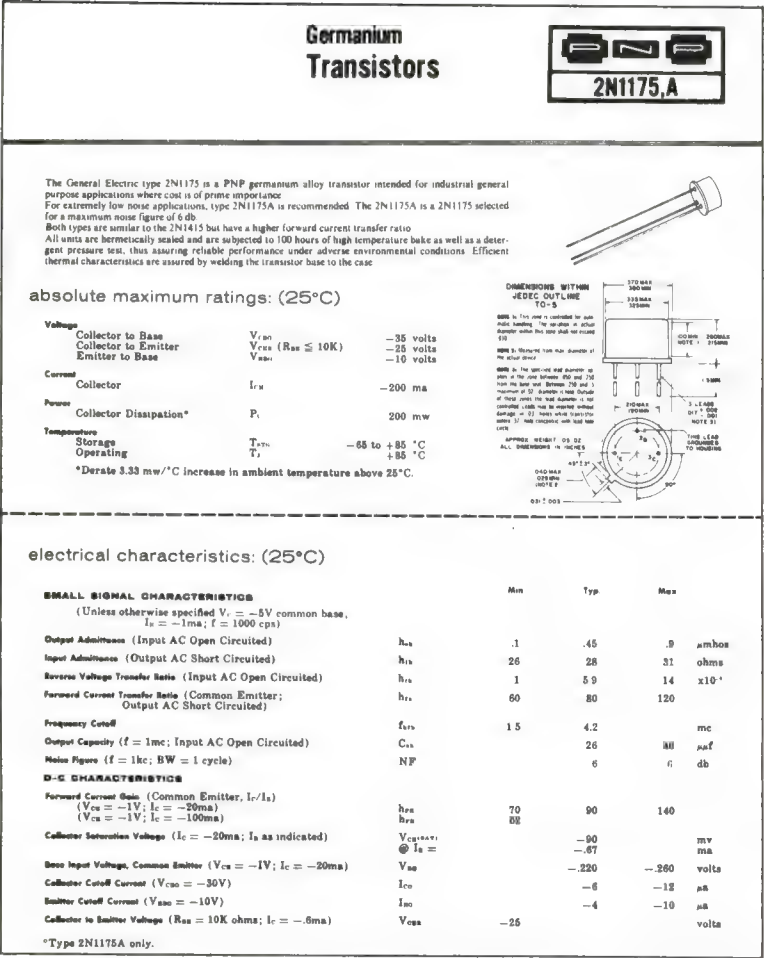


5. MAXIMUM SURGE CURRENT AT RATED LOAD CONDITIONS (NON-RECURRENT).



6. MAXIMUM SURGE CURRENT FOR SINGLE-CYCLE PULSE RATED LOAD CONDITIONS (NON-RECURRENT). CURVE USED IN CALCULATION OF  $I_T$

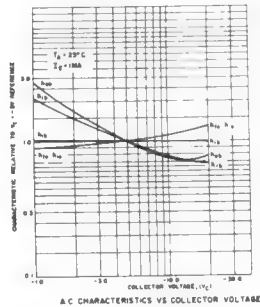
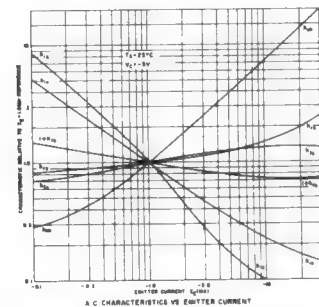
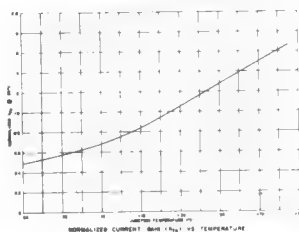
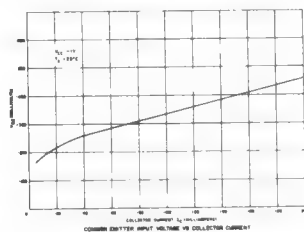
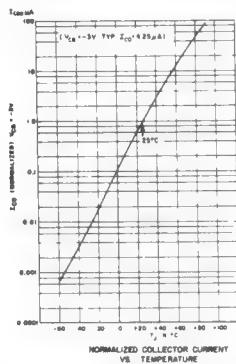
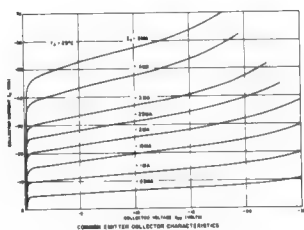
**Fig. 8.64 (Continued.)**



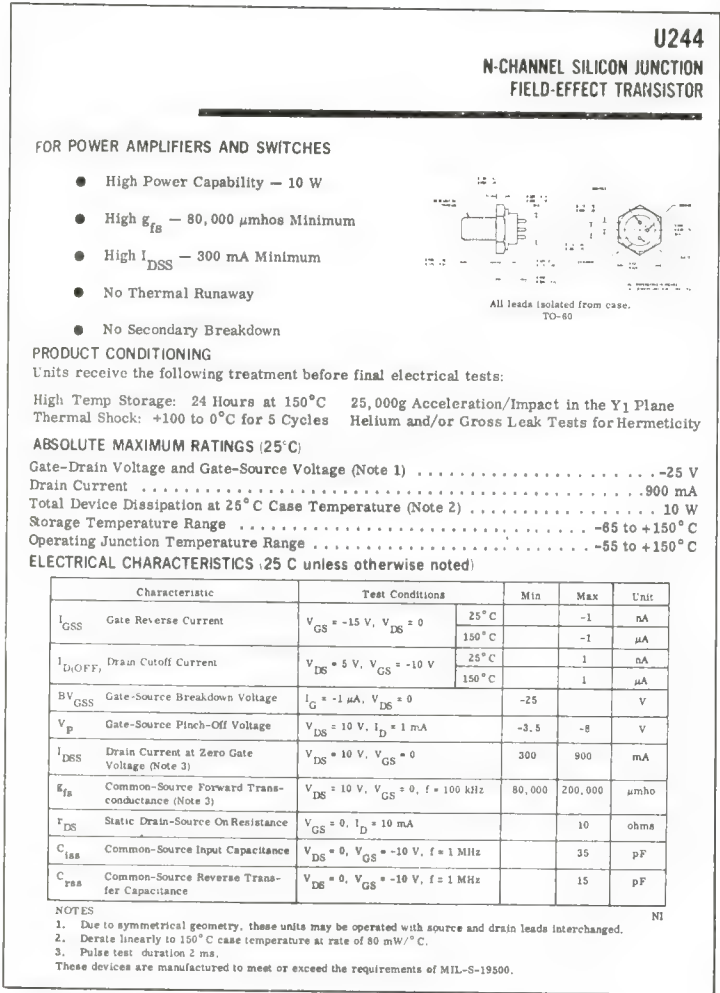
**Fig. 8.65** Typical data sheets for a bipolar junction transistor. (Courtesy General Electric Company.)



**2N1175,A**



**Fig. 8.65 (Continued.)**



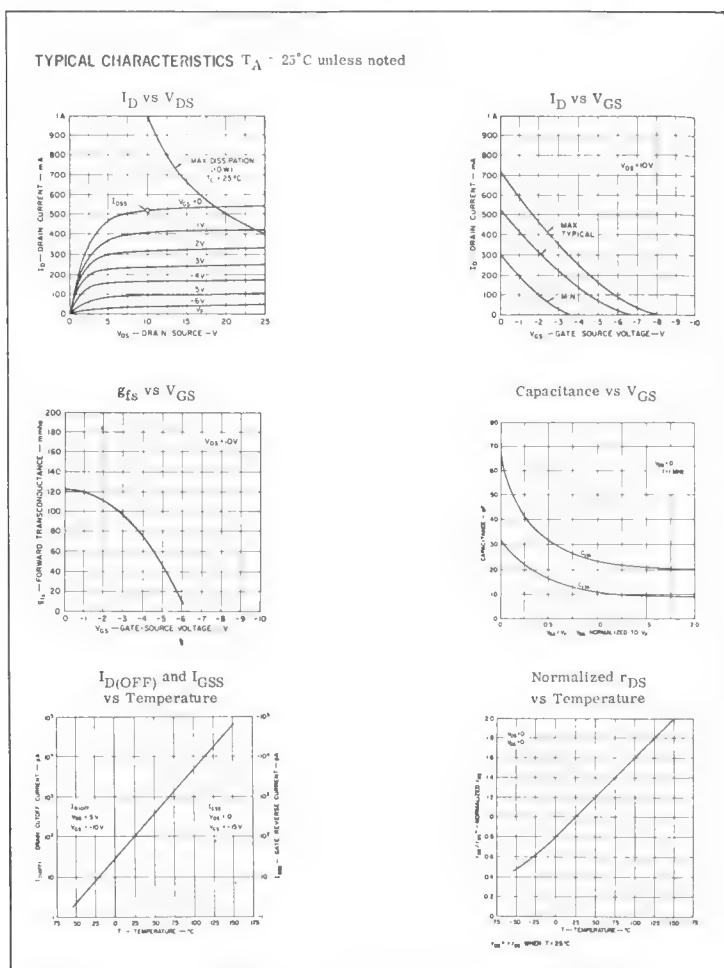


Fig. 8.66 (Continued.)

purpose applications where cost is of prime importance." The JFET is designed for power amplifier and switching applications.

2. A sketch with dimensions of the device: identification of connecting leads.
3. Maximum ratings, such as voltage, current, and temperature, that should never be exceeded.
4. Characteristics curves, such as the forward and reverse characteristics for a diode, collector characteristics for a BJT, and drain characteristics for a JFET.
5. Various curves of interest as a function of temperature, voltage, current, or frequency. For example, power derating and surge current curves are included for the diode. Interesting curves included for the BJT are the variation of the small-signal hybrid parameters with emitter current and collector voltage. The transfer characteristic ( $I_D$  versus  $V_{GS}$ ) is of interest for the JFET.
6. The spread of parameters by denoting their minimum, typical, and maximum values.

In making a selection, the data sheets for the device should be carefully studied and understood. Test condition for each parameter should be stated on the sheets. If data that are important for the intended application are missing, the data should be obtained.

## Chapter 9

# Integrated Circuit Technology

### 9.1 INTRODUCTION

Since the appearance of the integrated circuit (IC) in the early 1960s, the electronics industry has experienced monumental changes in design philosophy and manufacturing methods. This has inspired a host of new products, especially in the consumer sector, such as hand-held electronic calculators and electronic watches. In the future, besides innovations in the industrial market, many other products incorporating IC's will be developed for consumer use.

The IC evolved from attempts by semiconductor manufacturers to improve transistor performance and reduce costs. These efforts culminated in the late 1950s in the development of the planar diffused transistor. The techniques used in processing this device are essentially the same as those employed in making the most widely produced integrated circuit, the *monolithic* IC. Monolithic is derived from a Greek word meaning "single stone." In the monolithic IC, all components of the integrated circuit are formed in a single chip of silicon.

A family tree of integrated circuits manufactured is shown in Fig. 9.1. In addition to the monolithic IC, there is the *hybrid* which is available in two forms: *thin* and *thick film*. In either one, resistors, capacitors, and conductors are deposited on a substrate, such as ceramic. In the thin-film process, resistors, capacitors, and conductors are deposited on the substrate by evaporation of a suitable material in a vacuum. In the thick-film process, the components may be thought of as being "painted on" the substrate. Thin- and thick-film, as well as monolithic technology, will be described in this chapter.

Because of the initial high "tooling up" costs, monolithic technology is suited for mass production of IC devices. For small to moderate requirements, hybrid processing is profitable. If, for technical reasons, it is not feasible to develop a monolithic IC for a given circuit, the hybrid is made, regardless of quantity.

### 9.2 DIMENSIONING INTEGRATED CIRCUITS

A number of different dimensional units are utilized in defining the geometry of integrated circuits and their components. A comparison of these units is given in Table 9.1. Generally, *microns* are used for IC device dimensions and *mils* for the overall chip.

### 9.3 MONOLITHIC TECHNOLOGY

In the manufacture of IC's, hundreds to thousands of circuits are *processed simultaneously*. For this reason, the cost per circuit is low, making it possible for many IC am-

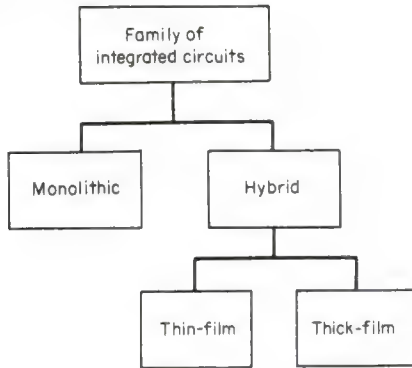


Fig. 9.1 A family tree of integrated circuits.

TABLE 9.1 Comparison of Dimensional Units for Integrated Circuits

	Mil	Centimeter	Micron	Angstrom
Mil	1	$2.54 \times 10^{-3}$	25.4	$2.54 \times 10^3$
Centimeter, cm	$3.94 \times 10^2$	1	$10^4$	$10^8$
Micron, $\mu$	$3.94 \times 10^{-2}$	$10^{-4}$	1	$10^{-4}$
Angstrom, $\text{\AA}$	$3.94 \times 10^{-6}$	$10^{-8}$	$10^{-4}$	1

For example, 1 mil =  $2.54 \times 10^{-3}$  cm = 25.4  $\mu$ ; 1  $\mu$  =  $10^{-4}$  cm; and 1  $\text{\AA}$  =  $10^{-4}$   $\mu$ .

plifiers to sell for less than a single transistor sold for a few years ago. The basic silicon material used in monolithic processing comes in a wafer, 1.5 to 3 in diameter and 6 mils thick, as shown in Fig. 9.2. The flat is used for reference, ensuring that the proper orientation of the wafer is maintained during the different processing steps.

Each of the squares in Fig. 9.2 is a chip that contains a complete integrated circuit. A typical size of a chip is  $\frac{1}{16}$  by  $\frac{1}{16}$  of an inch. Upon completion of processing, the wafer is *scribed* and then *fractured* along the vertical and horizontal lines. The chip is tested and packaged. Because of defects and loss of material along the periphery of the wafer, the number of good chips is less than the number theoretically possible from the wafer. The ratio of good chips to the maximum number of chips available from a wafer is referred to as the *yield* of the process.

**BASIC PROCESSING STEPS** To learn how a circuit is integrated, we shall first describe the processing steps required in integrating a single transistor (Fig. 9.3). Once

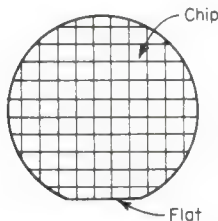


Fig. 9.2 A silicon wafer. Each chip contains a complete integrated circuit. The flat is used as a reference in the processing sequence of the wafer.

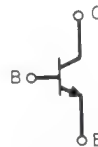


Fig. 9.3 A npn transistor to be integrated.

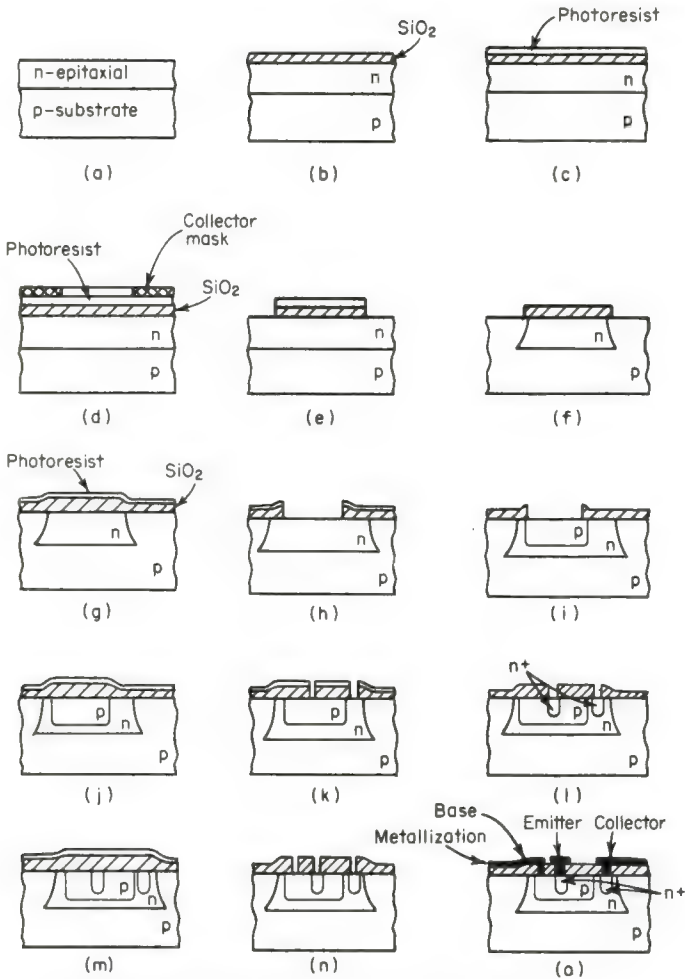


Fig. 9.4 Processing steps for integrating a transistor. (See text.)

the processing steps are understood, we shall see how an amplifier is integrated. Reference will be made to Fig. 9.4 in the following discussion.

**P-type substrate** The starting material is generally p-type silicon having a resistivity of about  $10 \Omega\text{-cm}$ . The material has a *single-crystal* structure. Single crystal denotes a material in which the atoms are arranged in an ordered pattern. Material in which the atoms are not in an ordered pattern is referred to as *polycrystalline*.

**Epitaxial layer** Upon the p-type silicon a well-defined n-type silicon region is grown in an epitaxial furnace. *Epitaxial growth* means that the n region formed has the same single-crystal structure as the p-type silicon (Fig. 9.4a).

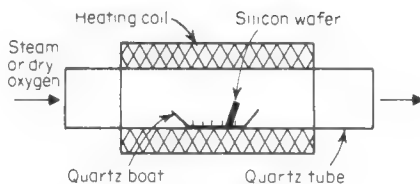
**Oxide layer** The silicon wafer, seated in a *quartz boat*, is placed in an *oxidation furnace* of Fig. 9.5 and exposed to steam or dry oxygen at a temperature between  $1000$  and  $1200^\circ\text{C}$ . The surface of the silicon is oxidized, and a thin layer of silicon dioxide ( $\text{SiO}_2$ ), which is glass, is formed (Fig. 9.4b).

An oxide layer has two important functions:

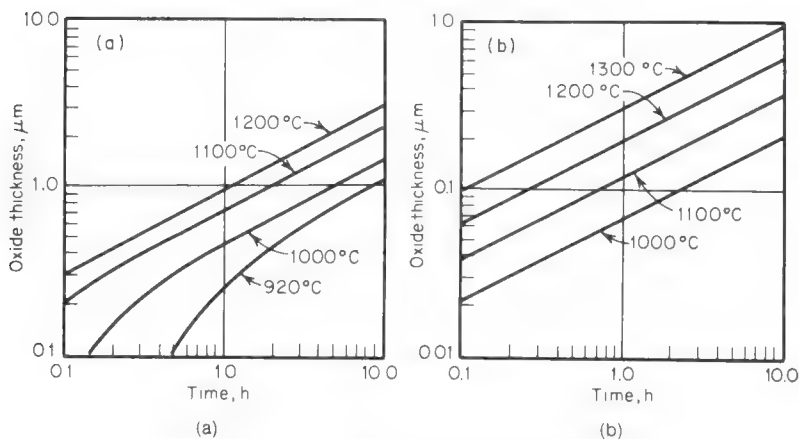
1. It serves as the surface for a light-sensitive film.
2. It protects the circuit against contamination, a serious problem in processing integrated circuits. This property is referred to as *passivation*.



## 9.4 Integrated Circuit Technology



**Fig. 9.5** Cross-sectional view of a basic oxidation furnace.



**Fig. 9.6** Oxidation curves: (a) Steam. (b) Dry oxygen.

Curves of  $\text{SiO}_2$  growth in steam and dry oxygen as a function of time for different temperatures are illustrated in Fig. 9.6. The application of the curves is illustrated in our first example.

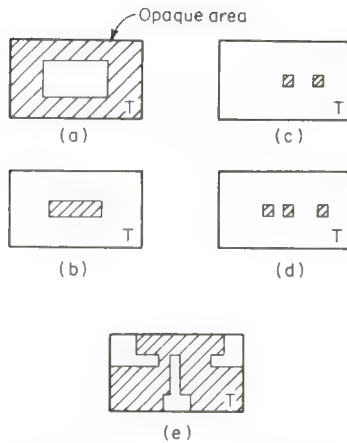
**example 9.1** Using the oxidation curves of Fig. 9.6, estimate the time needed to grow a  $0.2\text{-}\mu$  thickness of  $\text{SiO}_2$  on silicon, using steam and dry oxygen. The temperature of the oxidation furnace is  $1100^\circ\text{C}$ .

**solution** It is noted that the vertical and horizontal scales of the oxidation curves are not linear. To cover a large range of values ( $10/0.1 = 100:1$ ) in a compact manner, *logarithmic scales* are used. Logarithmic scales in electronics are nothing new. For example, the frequency scale for the frequency-response curve of an amplifier is also logarithmic. Referring to the  $1100^\circ\text{C}$  steam curve of Fig. 9.6a, for an oxide thickness of  $0.2\text{ }\mu$ ,  $0.1\text{ h}$  (6 min) is required. If dry oxygen is used, from the  $1100^\circ\text{C}$  curve of Fig. 9.6b,  $3\text{ h}$  (180 min) is necessary. Because it takes less time to grow an oxide layer, steam is generally used.

**Photoresist** An organic substance that is sensitive to ultraviolet (UV) light, called *photoresist*, is applied over the oxide layer (Fig. 9.4c). A commonly used resist is *Kodak Photo Resist* (KPR). The wafer is held by vacuum on a *spinner*. Some drops of KPR are applied to the surface, and the wafer is spun at a few thousand rpm for 10 to 30 s. This ensures a uniform deposition of KPR over the wafer's surface.

Upon exposure to UV, the photoresist becomes *polymerized*. When this occurs, the molecular structure of the exposed resist is altered. The change results in the exposed material becoming *resistant* to chemicals used in subsequent processing such as hydrofluoric acid (HF) for etching and phosphorus (P) and boron (B) for doping. The unexposed areas, however, are affected by these materials.

**Artwork** Each chip on the wafer of Fig. 9.2, which may measure  $\frac{1}{16}$  by  $\frac{1}{16}$  in or less, contains perhaps a dozen or more transistors and an equal number of resistors. It becomes apparent that the regions on a chip must be precisely defined. For an npn transistor, three regions are needed: collector, base, and emitter. The three regions are cut as masks, illustrated in Fig. 9.7. A Mylar base, which is dimensionally stable with



**Fig. 9.7** Set of five masks required for integrating the transistor of Fig. 9.3. (See text.)

temperature and humidity, on which is a peelable red tinted layer, is used for the mask material. Commercially the material is referred to as *Rubylith*.

The pattern is formed by cutting through the tinted layer (and not the Mylar base) and peeling away the unwanted layer. To achieve the required precision, the pattern is cut on a mask of the size in the order of 20 by 30 in. This is accomplished on a *co-ordinatograph* (Fig. 9.8), a machine that provides the required cutting accuracy. The tee in each mask is used in mask alignment (to be described later).

The mask is ultimately reduced to the actual size of the chip (for example,  $\frac{1}{16}$  by  $\frac{1}{16}$ -in chip size). In addition, the pattern must be repeated to cover the whole wafer. This is done in a *step-and-repeat camera*, which reduces the mask to its final size and repeats the pattern on a high-resolution emulsion glass plate. A plate is made for each of the masks.

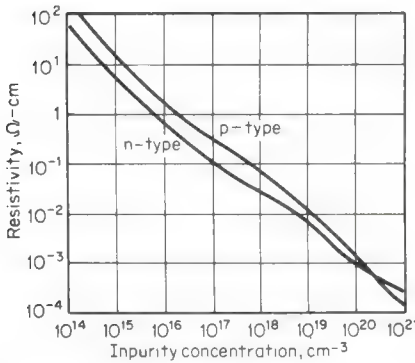


**Fig. 9.8** A coordinatograph is used for the precision cutting of masks. (Courtesy Owen Jones Equipment Sales.)

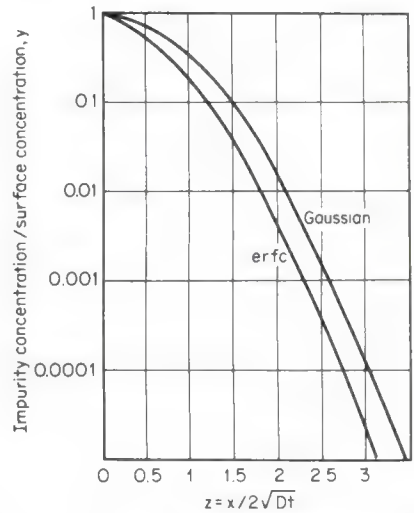
**Exposure of collector mask** The cross-hatching in Fig. 9.7 denotes the opaque areas of the mask. When exposed, the UV is prevented by the opaque areas on the mask from reaching the resist; the UV, however, strikes the resist through the clear mask areas. The collector mask superimposed on the wafer is shown in Fig. 9.4d.

**Diffusion** The key to monolithic IC technology is the diffusion of impurities, called *dopants*, into silicon. In diffusion, p-type such as boron (B), and n-type, such as phosphorus (P), dopants in high concentration are brought in contact with the silicon surface. The process occurs in a diffusion furnace, which is similar in basic construction to the oxidation furnace of Fig. 9.5. Diffusion occurs in the temperature range of 1100 to 1200°C. The dopant is typically in vapor or gaseous form. The penetration depth and concentration of impurity in the silicon depend on the type of dopant, time, and temperature of diffusion, and on how the dopant is introduced.

Silicon, like other materials, can absorb only a finite number of impurity atoms, called the *solid solubility* of the material. Solid solubility is the maximum number of atoms of one material that can be dissolved in another material.



**Fig. 9.9** Curves of silicon resistivity for p- and n-type impurity concentrations at room temperature (25°C).



**Fig. 9.10** Erfc and gaussian distribution curves.

Curves of silicon resistivity (ohm-centimeters) for p- and n-type *impurity concentrations* (impurity atoms per cubic centimeter of silicon) are given in Fig. 9.9. It is seen that, as the impurity concentration is increased, the resistivity of the silicon is decreased.

There are two methods for applying dopants to the wafer in the diffusion furnace. In the first method, the wafer *always* “sees” a *constant* concentration,  $N_0$  atoms/cm<sup>3</sup>, of dopant. The resulting distribution of the impurity in silicon is described by what is called the complementary error function (erfc) curve of Fig. 9.10. In the second method, the wafer sees a *fixed* amount of dopant atoms. The resulting distribution is gaussian and is also shown in Fig. 9.10.

For determining the location of a pn junction in silicon, Fig. 9.10 is extremely useful. Both curves are plotted as  $N_x/N_0$  versus  $z = x/2\sqrt{Dt}$ . Term  $N_x$  is the actual concentration of impurity atoms per cubic centimeter at a distance  $x$  from the surface of the silicon;  $N_0$  has already been defined. Term  $D$  is the diffusion coefficient in square centimeters per second; it indicates how fast the dopant spreads out in the silicon. Plots of  $D$  for different dopants as a function of temperature are given in Fig. 9.11.

**example 9.2** A uniformly doped p-type silicon wafer is used in the manufacture of an integrated circuit. The resistivity of the wafer measures 10 Ω-cm. It is placed in a diffusion

furnace at 1200°C and is exposed to a *constant source* of phosphorus,  $N_0 = 10^{19}$  atoms/cm<sup>3</sup>. How long does it take for a pn junction to form 3  $\mu$  away from the surface?

**solution** Referring to the resistivity curves of Fig. 9.9, for p-type silicon of 10  $\Omega$ -cm resistivity, the impurity concentration is about  $1.4 \times 10^{15}$  atoms/cm<sup>3</sup>. This figure is called the *background concentration*,  $N_B$ , of impurity. In this example, then,  $N_B = 1.4 \times 10^{15}$  atoms/cm<sup>3</sup>.

A constant source of dopant (phosphorus) implies a complementary error function (erfc) distribution. A pn junction is formed when the concentration of the n-type dopant (phosphorus) equals the p-type dopant (background concentration). Hence,  $N_x = N_B = 1.4 \times 10^{15}$ .

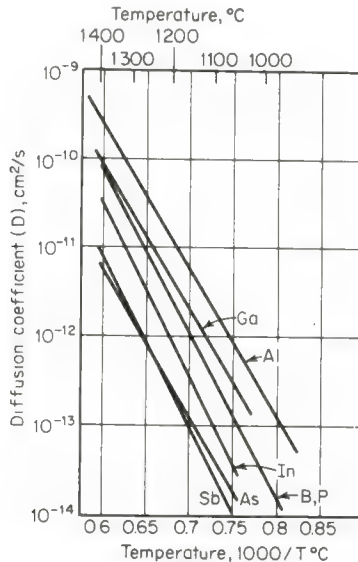


Fig. 9.11 Diffusion coefficients for various impurities in silicon.

The ratio  $N_x/N_0 = 1.4 \times 10^{15}/10^{19} = 1.4 \times 10^{-4} = 0.00014$ . From the erfc curve of Fig. 9.10, it is estimated that  $z = 2.6$ .

At 1200°C, the diffusion coefficient for phosphorus, from Fig. 9.11, is  $D_p \approx 3 \times 10^{-12}$  cm<sup>2</sup>/s. The junction depth  $x$  is 3  $\mu = 3 \times 10^{-4}$  cm (see Table 9.1). Substitution of these values in  $z$  yields

$$2.6 = \frac{3 \times 10^{-4}}{2\sqrt{3 \times 10^{-12}t}}$$

Squaring both sides of the equation, we obtain

$$6.75 = \frac{9 \times 10^{-8}}{4 \times 3 \times 10^{-12}t}$$

Solving for  $t$ ,

$$t = \frac{9 \times 10^{-8}}{4 \times 3 \times 10^{-12} \times 6.75} \\ \approx 1100 \text{ s} \approx 0.3 \text{ h}$$

**Etching** The unexposed SiO<sub>2</sub> is etched away by hydrofluoric acid (HF) (Fig. 9.4e).

**p-type diffusion** The wafer is placed in a diffusion furnace, and sufficient p-type impurity, such as boron, is introduced. The impurity penetrates the silicon which was etched away by HF. After a length of time, an n region (or *island, tub*) for the collector is formed (Fig. 9.4f). Because the impurity diffuses *horizontally* (laterally), as well as vertically, the profile of Fig. 9.4f is the result. Note that some SiO<sub>2</sub> overlaps the n region. This is desirable because it helps to minimize contamination. The photoresist is removed, or *stripped*, after each diffusion.

**Oxidation and resist** The wafer is placed again in the oxidation furnace, and a new layer of  $\text{SiO}_2$  is grown. Photoresist is also applied (Fig. 9.4g). The wafer is now ready for exposure to the base mask.

**Exposure to base mask** The base mask of Fig. 9.7b is now exposed to UV. The unexposed regions are etched away, and the result is an opening of a "window" for base diffusion (Fig. 9.4h).

To ensure perfect alignment between the base mask and the wafer, a *mask aligner* is used. With this equipment, the operator looks through a microscope and adjusts the position of the mask so that the tee of the base mask aligns with the tee on the wafer from its exposure to the collector mask.

**Base diffusion** A p-type impurity is diffused through the window to form the base of the transistor (Fig. 9.4i).

**Oxidation and resist** A new layer of oxide is grown and photoresist applied (Fig. 9.4j).

**Forming the emitter** The emitter is formed by using a very high concentration of an n-type dopant, denoted by  $n+$ . (If  $p+$  was indicated, this would mean a very high concentration of a p-type dopant.) Further, to ensure a *nonrectifying* or, *ohmic, contact* to the n-collector region,  $n+$  doping is also used.

The connections between the diffused regions in a monolithic IC are generally made by a thin film of conducting material, such as aluminum. Aluminum is a p-type material. If a p material is in contact with an n material, a pn junction diode is formed. An  $n+$  region in contact with aluminum, however, provides a nonrectifying contact.

**Exposure to  $n+$  mask** The mask of Fig. 9.7c has cutouts for the emitter region and for a contact to the collector region. The mask is exposed to UV, and, as in previous processing steps, the unexposed regions are etched away (Fig. 9.4k).

**$n+$  diffusion** A high concentration of n-type dopant, such as phosphorus, is diffused through the windows to form the  $n+$  regions (Fig. 9.4l).

**Oxidation and resist** A new layer of oxide is grown and photoresist applied (Fig. 9.4m).

**Exposure to contact mask** Windows must now be opened to permit electric connections to be made to the emitter, base, and collector regions of the transistor. The contact mask of Fig. 9.7d is exposed to UV. The unexposed regions are etched away, and the resist is stripped from the wafer (Fig. 9.4n).

**Metallization** To provide connections between the diffused regions, a metallic



Fig. 9.12 An example of a prober. (Courtesy Electroglas, Inc.)



material is deposited over the entire wafer. This process is called *metallization*. Typically, aluminum is used and is deposited as a thin film on the substrate, which is accomplished in a vacuum. (Vacuum deposition of materials is considered later in the chapter.)

**Exposure to interconnection mask** The deposited aluminum is covered with photoresist and exposed to the interconnection mask of Fig. 9.7e. The unexposed regions are etched away, and the resist is removed from the wafer. The result is a diffused (integrated) transistor (Fig. 9.4o).

**Testing and packaging** The integrated circuit is tested and packaged. In testing, a *prober* (Fig. 9.12) is used. The wafer is vacuum-held and, under a microscope, fine points are brought into contact with the diffused regions. Through external electric connections to the points, voltages are impressed across the regions and measurements of device performance are made.

In summary, five different masks and three diffusions were required in the fabrication of a monolithic transistor. As we shall see, the same number of masks and diffusions are all that is generally required in the monolithic integration of a complete circuit. This feature is what makes monolithic technology so attractive for large-volume production of integrated circuits.

## 9.4 ELECTRICAL ISOLATION

In the monolithic IC all circuit elements, such as transistors and resistors, share the same silicon chip. It is necessary, therefore, that the elements be *electrically isolated* from one another. At present, three methods are used for isolation:

1. Junction
2. Dielectric
3. Beam lead

**JUNCTION ISOLATION** We have seen how a single transistor is integrated. To illustrate junction isolation, consider the two transistors in a silicon chip shown in Fig. 9.13. Note that the p substrate is returned to a negative potential and the n regions to a positive potential. Because the n and p regions form pn junctions, they are *reverse-biased*, similar to a back-biased semiconductor diode. For this condition, no current flows between the n-collector regions through the p substrate, and electrical isolation is obtained.

**DIELECTRIC ISOLATION** An example of two transistors dielectrically isolated is illustrated in Fig. 9.14. Surrounding each n-collector region and, separating it from the substrate, is a layer of silicon dioxide. Because  $\text{SiO}_2$  is an insulator, electrical isolation is realized. The substrate does not have to be single-crystal silicon; instead, polycrystalline silicon is used.

**BEAM-LEAD ISOLATION** In this method of isolation, the circuit elements are diffused as in the junction-isolated IC. The metallization deposited on the chip, however, is very thick (in the order of 1 mil). The superfluous silicon of the p substrate is re-

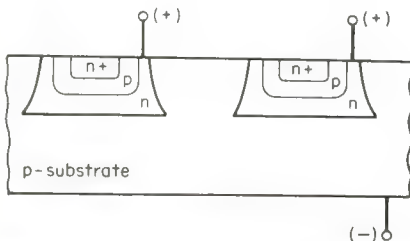


Fig. 9.13 Junction isolation for two diffused transistors.

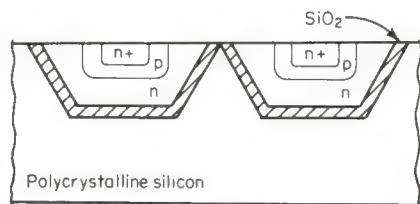


Fig. 9.14 Dielectric isolation for two diffused transistors.



moved. The resulting structure is a semirigid connected circuit with all elements physically separated from one another.

Owing to lower production costs, most monolithic circuits use junction isolation.

**PARASITICS** A problem with junction isolation is the existence of parasitics. A parasitic may be defined as an *undesired element*. Referring to Fig. 9.13, we saw that

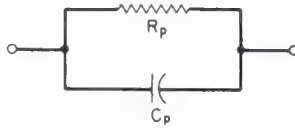


Fig. 9.15 Parasitic elements existing across a reverse-biased pn junction.

isolation was obtained by reverse-biasing the n-collector and p-substrate regions. In reverse-biasing a pn junction, a capacitance  $C_p$ , in parallel with a high-valued resistance  $R_p$ , exist as illustrated in Fig. 9.15. These two elements are examples of parasitics and, at very high frequencies, may create problems. Parasitics also exist for diffused resistors and capacitors in the junction-isolated IC.

## 9.5 PROPERTIES AND CHARACTERISTICS OF DIFFUSED ELEMENTS

**BIPOLAR JUNCTION TRANSISTOR (BJT)** A diffused BJT has typical values of short-circuit current gain  $h_{fe}$  as high as 300 and gain-bandwidth product  $f_T$  as high as 500 MHz. Because the gain of a junction transistor depends on the width of its base, the short-circuit current gain can be increased by reducing the base thickness. Transistors with very narrow base widths, exhibiting short-circuit current gains as high as 5000, are called *super-beta* and *punch-through* transistors. Owing to the narrow base width, the collector-emitter voltage for this type of transistor is limited to about 0.5 V. For a conventional BJT, the collector-emitter voltage is in the order of 20 to 30 V.

The gain-bandwidth product may be increased by decreasing the resistance of the collector region. A method for realizing this, without affecting appreciably the collector-emitter voltage rating of the transistor, is shown in Fig. 9.16. A heavily doped region, called an *n<sup>+</sup> buried layer*, is diffused between the bottom of the n-collector region and the p substrate. Because the n<sup>+</sup> layer is in parallel with the n-collector region, the effective collector resistance is reduced.

**LATERAL pnp TRANSISTOR** The diffusion of npn transistors is very compatible with the monolithic process. There are occasions, however, when a pnp transistor is needed in addition to an npn device. A structure that requires no additional diffusion steps is the *lateral pnp transistor* of Fig. 9.17.

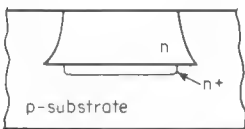


Fig. 9.16 An n<sup>+</sup> buried layer for reducing collector resistance.

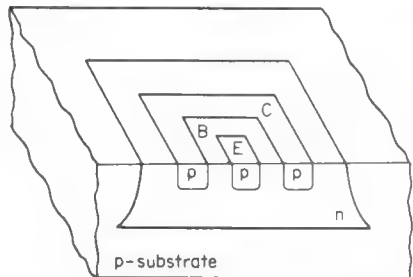


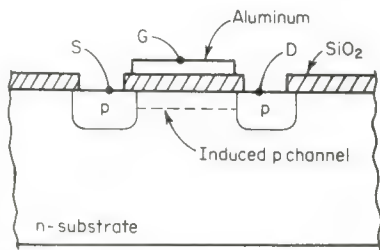
Fig. 9.17 A view of a lateral pnp transistor.

Two p-type regions are diffused in the n island (tub). One of the p regions serves as the emitter and the other as the collector. Because the effective base width is greater than that of an npn transistor, the short-circuit current gain and gain-bandwidth product of a lateral pnp transistor are appreciably less. Typical maximum values are  $h_{fe} = 20$  and  $f_T = 2$  MHz.

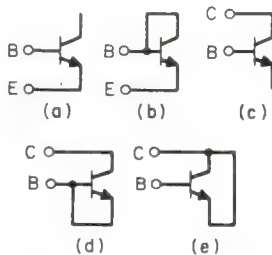
**FIELD-EFFECT TRANSISTOR (FET)** In the integration of field-effect transistors, the enhancement-mode MOSFET (IGFET) is invariably chosen because

1. It is self-isolating.
2. It can be formed by a single diffusion.
3. Because it is self-isolating, it has a much greater packing density than the BJT.

A cross-sectional view of a p-channel enhancement-mode MOSFET is shown in Fig. 9.18. Two p-type regions are diffused simultaneously in the n substrate. The gate is formed by depositing aluminum over the oxide layer between the p regions. In operation, the substrate is returned to the most positive potential in the circuit. The source, gate, and drain are so biased that they are negative with respect to the substrate. Thus, reverse-biased pn junctions are formed and electrical isolation is achieved.



**Fig. 9.18** A cross-sectional view of a p-channel enhancement MOSFET.



**Fig. 9.19** A transistor may be connected in five ways to act like a diode (see text).

**DIODE** By utilizing only two terminals of a diffused npn transistor, five different diode configurations may be realized. Referring to Fig. 9.19, these are:

- a. Base-emitter junction diode; collector is open. Series diode resistance and reverse breakdown voltage are low.
- b. Base-emitter junction diode; collector short-circuited to base. Series diode resistance and reverse breakdown voltage are low.
- c. Base-collector junction diode; emitter open. Series diode resistance and reverse breakdown voltage are high.
- d. Base-collector junction diode; emitter short-circuited to base. Series diode resistance and reverse breakdown voltage are high.
- e. Base-collector junction diode; emitter short-circuited to collector. Series diode resistance is high, and reverse breakdown voltage is low.

A series diode resistance which is low is less than  $50 \Omega$ ; a high series diode resistance is greater than  $50 \Omega$ . Low reverse breakdown voltage is in the order of 7 V; high reverse breakdown voltage is about 40 V. Because of its good properties, the diode of Fig. 9.19b is commonly used.

**RESISTOR** Integrated resistors are generally formed by the diffusion of p-type material in an n tub, as illustrated in Fig. 9.20. This diffusion is done *simultaneously* with the diffusion of the p-base region of an npn transistor.

Resistance,  $R$  (ohms), of a material is expressed by

$$R = \rho \left( \frac{L}{A} \right) \quad (9.1)$$

where  $\rho$  = resistivity of material,  $\Omega$ -cm;  $L$  = length of material, cm; and  $A$  = cross-

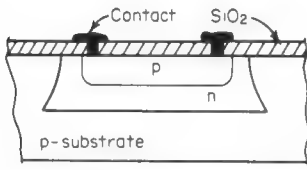


Fig. 9.20 A cross-sectional view of a p-diffused resistor.

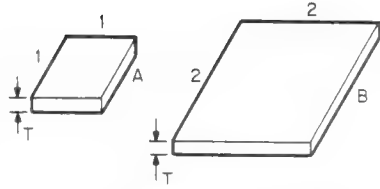


Fig. 9.21 The resistance measured across any two parallel edges is the same for either square.

sectional area of material,  $\text{cm}^2$ . The area may be expressed by the product of its width  $W$  and its thickness,  $T$ :  $A = WT$ . Substitution of this value for  $A$  in Eq. (9.1) yields

$$R = \frac{\rho}{T} \frac{L}{W} \quad (9.2)$$

If  $L = W$ , we have a *square* of material, and Eq. (9.2) reduces to

$$R = \frac{\rho}{T} \quad (9.3)$$

Assuming that thickness  $T$  is constant, which is reasonable, then the ratio  $\rho/T$  is also constant. Letting this quantity be denoted by  $R_s$ , the *sheet resistance*, we obtain

$$R = R_s \text{ ohms/square} \quad (9.4)$$

For any length and width of a diffused resistor,

$$R = R_s \left( \frac{L}{W} \right) \quad (9.5)$$

where  $L/W$  is the *aspect ratio*.

Consider the two square areas of Fig. 9.21. Both squares have the same thickness  $T$ . Although the surface area of square  $B$  is four times as large as that of square  $A$ , the resistance measures across their edges are the same and equal to the sheet resistance of the material  $R_s$ .

The last statement can be verified by noting that for square  $A$ , the cross-sectional area is  $T(1) = T$ . For square  $B$ , the area is  $T(2) = 2T$ . The length of  $A$  is 1 unit, and of  $B$ , 2 units. Hence, the length to cross-sectional areas are equal:  $1/T = 2/2T = 1/T$ . The resistance for each square is identical.

**example 9.3** For a diffused resistor of  $2000 \, \Omega$ ,  $R_s = 200 \, \Omega/\text{square}$ . Determine the aspect ratio and, if  $L = 20 \, \mu$ , the width of the resistor.

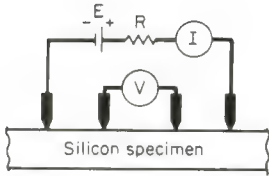
**solution** From Eq. (9.5),  $R/R_s = L/W = 2000/200 = 10:1$ . If  $L = 20 \, \mu$ ,  $W = 20/10 = 2 \, \mu$ .

The practical range of  $p$ -diffused resistors is from  $20 \, \Omega$  to  $30 \, \text{k}\Omega$ . Their absolute tolerance, however, is poor, being in the order of  $\pm 25$  percent. The tolerance of the ratio of two diffused resistors can be as low as  $\pm 2$  percent. Resistors made with  $n$ -diffusion are seldom used.

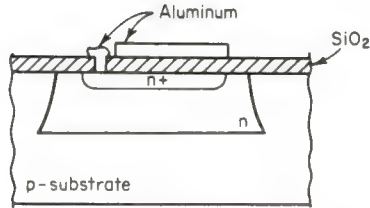
**FOUR-POINT PROBE** The four-point probe of Fig. 9.22 permits a simple measurement of the sheet resistance. A dc voltage source in series with a resistor, which simulates a *current source*, is connected to the outer points in series with a dc ammeter,  $I$ . The points are spaced about  $0.1 \, \text{cm}$  apart and make physical contact with a silicon specimen whose sheet resistance is to be determined. The inner points are connected to a dc voltmeter which reads the voltage  $V$  developed across the points. Sheet resistance, in ohms per square, is given by

$$R_s = 4.5 \frac{V}{I} \quad (9.6)$$

**CAPACITOR** Capacitors found in a monolithic IC are basically of two kinds: *junction* and *thin film*. A junction capacitor may be realized by using an npn transistor



**Fig. 9.22** A four-point probe used for measuring sheet resistance.



**Fig. 9.23** A cross-sectional view of a thin-film monolithic capacitor.

connected as a diode (Fig. 9.19). Application of a reverse voltage across any two terminals results in a variation in capacitance: the greater the reverse voltage, the less is the capacitance.

A cross-section view of a thin-film monolithic capacitor is shown in Fig. 9.23. An  $n$ -region, diffused in the  $n$  tub, acts as one plate of the capacitor. A growth of silicon dioxide serves as the dielectric, and the deposition of aluminum over the oxide acts as the other plate of the capacitor.

Capacitors occupy a relatively large area on a chip and are limited to small values of capacitance, typically 200 pF. For these reasons, then, they are avoided wherever possible.

**INDUCTOR** Currently there are no practical methods for diffusing inductors in a monolithic IC. Inductors, therefore, are avoided in the design of monolithic integrated circuits.

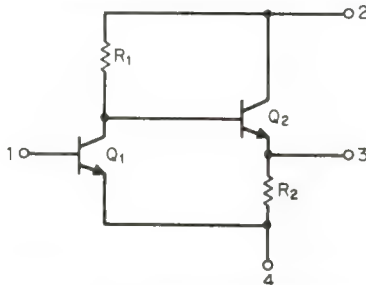
## 9.6 MAKING A MONOLITHIC IC

In this section we shall see how a circuit is transformed into a monolithic IC. Because it is most common, junction isolation will be assumed. In determining how many isolation tubs are required, the following guidelines are offered:

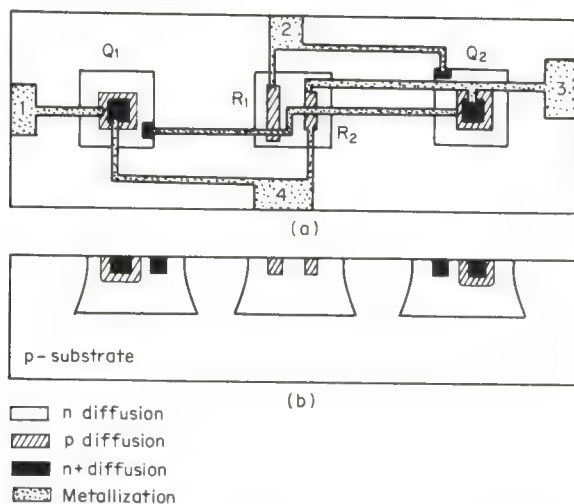
1. Transistor collectors at different potentials must be in separate isolation tubs.
2. Resistors connected to circuit points at different potentials can share the same isolation tub. The  $n$  tub should be connected to the most positive voltage available in the circuit.
3. Diffused capacitors generally require individual tubs.
4. Base-emitter diodes can generally share the same isolation tub.

Consider the two-stage direct-coupled amplifier of Fig. 9.24, which is to be integrated. Because the collector potentials of  $Q_1$  and  $Q_2$  are different, each transistor requires a separate isolation tub. Resistors  $R_1$  and  $R_2$  can share one tub. Therefore, three isolation regions are required.

A layout of the amplifier is given in Fig. 9.25a. Other layouts are also possible. A cross-sectional view of the integrated circuit is given in Fig. 9.25b. As in the integra-



**Fig. 9.24** A two-stage amplifier to be integrated.



**Fig. 9.25** Monolithic integration of the amplifier of Fig. 9.24:  
(a) Layout. (b) Cross-sectional view.

tion of a transistor, a p-type substrate with a grown n-epitaxial layer is the starting material. Also, five masks are required; referring to Fig. 9.26, these are:

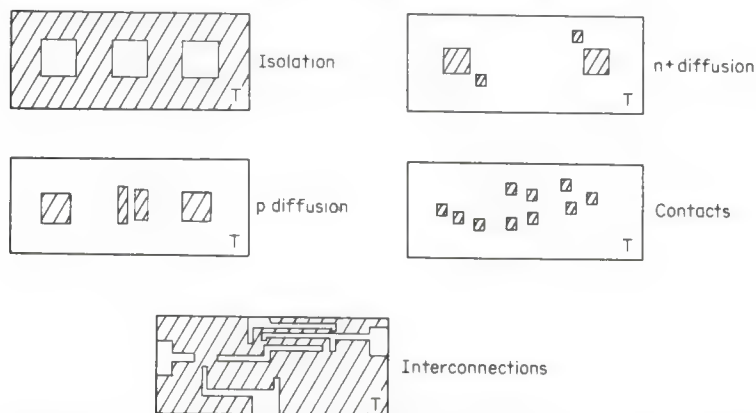
**Isolation mask** Sufficient p-type impurity is diffused in the n-epitaxial region to ensure that the three n-isolation tubs are surrounded by p regions (Fig. 9.25b). The opaque areas of the isolation mask prevent UV from reaching the photoresist on the silicon wafer; consequently, windows are only opened for the diffusion of the p-type impurity.

**p-diffusion mask** This mask has windows for the transistor bases and resistors. The diffusion of the base and resistors occurs simultaneously.

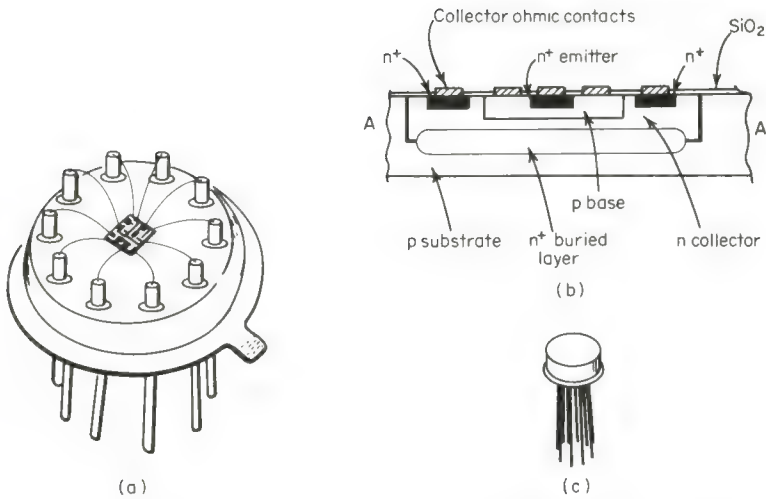
**n+ diffusion mask** Windows are cut out for the diffusion of emitters and regions in n-type material to ensure nonrectifying contacts.

**Contact and interconnection masks** The contact mask has cutouts for connections, and the interconnection mask has the connection pattern for the diffused elements.

As for the integration of a transistor, described earlier in the chapter, oxidation, application of photoresist, etc., are used in similar sequence for integration of a circuit. An example of a commercial monolithic IC is illustrated in Fig. 9.27.



**Fig. 9.26** Set of five masks required for the integration of the amplifier of Fig. 9.24 (see text).



**Fig. 9.27** An example of a commercial IC: (a) Constructional details. (b) Cross-sectional view of transistor. (c) Approximate size of package.

**BURIED CROSSOVER** In the layout of an integrated circuit it is found occasionally that two conducting paths, which cannot be eliminated, intersect each other. To correct this situation, a buried crossover, illustrated in Fig. 9.28, is required.

An  $n^+$  region is diffused in a separate  $n$ -isolation tub. Conductor 1 completes its path through the  $n^+$  buried region. Conductor 2, shown perpendicular to conductor 1, is insulated from the buried region by the  $\text{SiO}_2$  layer.

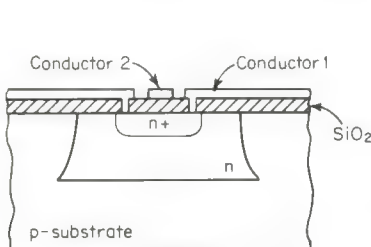
**example 9.4** The circuit of Fig. 9.29 is to be integrated. Draw a layout of the circuit, showing the required isolation tubs; the devices in each tub may be shown schematically. The numbering of terminals must be maintained in sequence on the periphery of the layout.

**solution** The layout, shown in Fig. 9.30, requires four isolation tubs. Because they share a common connection, the three diodes are placed in one tub. Owing to different collector potentials, each transistor requires its own isolation tub. The three resistors can share a single tub.

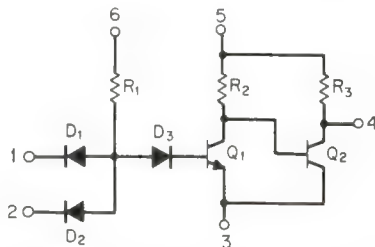
**example 9.5** A magnified drawing of an integrated-circuit layout is given in Fig. 9.31. Draw a schematic diagram of the circuit, and identify all components and the values of resistors. The sheet resistance  $R_s = 200 \Omega/\text{square}$ .

Assume that the transistor is saturated; that is,  $E_{CE} = 0 \text{ V}$ . If the voltage across a forward-biased base-emitter junction or diode is  $0.7 \text{ V}$ , determine the base, emitter, and collector currents in the transistor.

**solution** The complete circuit is illustrated in Fig. 9.32. Resistance values were obtained by counting the number of squares in each  $p$ -diffused resistor in the isolation tub. Diode  $D$  is always reverse-biased. The base current, therefore, is  $I_B = (6 - 0.7)/5.6 = 0.95 \text{ mA}$ . Col-

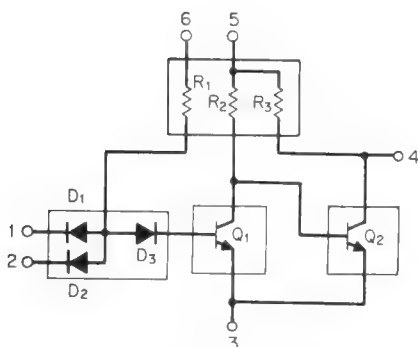


**Fig. 9.28** A buried crossover.

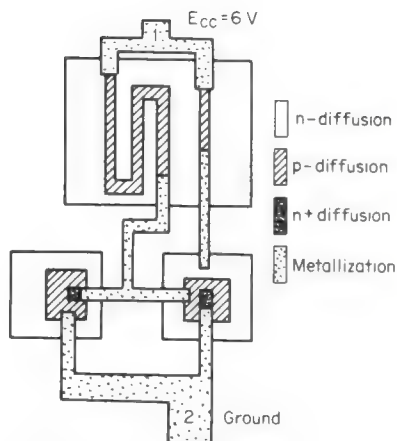


**Fig. 9.29** A circuit to be integrated. (See Example 9.4.)

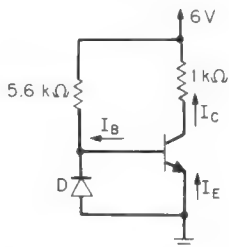




**Fig. 9.30** Schematic layout of the circuit of Fig. 9.29.



**Fig. 9.31** A given layout of an integrated circuit. (See Example 9.5.)



**Fig. 9.32** Schematic diagram of circuit determined from the IC layout of Fig. 9.31.

lector current,  $I_C = (6 - 0)/1 = 6$  mA. The emitter current  $I_E$  is equal to the sum of base and collector currents:  $I_E = I_B + I_C = 0.95 + 6 = 6.95$  mA.

## 9.7 LARGE-SCALE INTEGRATION (LSI)

Large-scale integration (LSI) is an extension of monolithic technology where 100 or more interconnected circuits are in a single chip of silicon. (Integrated circuits containing fewer than 100 interconnected circuits are often referred to as *medium-scale integration*, MSI.) Because of such a large number, invariably some diffused components are defective. What is necessary, therefore, is the elimination of the defective units before the interconnections are made. One commonly used technique for accomplishing this is called *discretionary wiring*.

In discretionary wiring, a number of circuits, called *unit cells*, in excess of the actual number required, are integrated. Generally, LSI is best suited where identical circuits are used, such as logic gates and computer memories. Each unit cell is checked on an automatic tester. Coupled to the tester is a computer which remembers the good cells. The computer generates an interconnection pattern that includes only good cells. Because the location of defective cells will be different for other chips of the same type, the testing and computer routine is repeated for each chip.

# Chapter 10

## Tuned Circuits

### 10.1 INTRODUCTION

In previous chapters, the subject of *reactance* was discussed. It was noted that inductors (*coils*) have a quality that is called *inductive reactance*, the magnitude of which varies with the applied frequency. It was also noted that capacitors have a quality called *capacitive reactance*, the magnitude of which also varies as a function of frequency.

Of particular importance is the fact that inductive reactance, which is generally considered as being *positive* reactance, is vectorially opposite to capacitive reactance, generally considered as being *negative* reactance. That is, when they are combined in a circuit, inductive reactance tends to *cancel* capacitive reactance. Also, inductive and capacitive reactance behave oppositely with respect to changes in applied frequency. For a fixed value of inductance, inductive reactance *increases* as the applied frequency is increased, but the capacitive reactance of a fixed value of capacitance *decreases* as the applied frequency is increased.

The reference to inductive reactance as being positive and capacitive reactance as being negative should not be misinterpreted. In an inductive circuit, the voltage leads the current and, in a capacitive circuit the voltage lags behind the current. Therefore, the effect of an inductor is opposite to the effect of a capacitor. As a convenient way of distinguishing between the effects of the two reactances, we call one positive and the other negative.

Another way of distinguishing between the effects of inductive reactance and capacitive reactance is to represent them graphically as shown in Fig. 10.1. This illustration shows that the effects of  $X_L$  and  $X_C$  are in opposite directions. If the values of  $X_L$  and  $X_C$  are equal, their effects cancel, leaving only resistance in the circuit. If the three quantities shown in Fig. 10.1 ( $X_L$ ,  $X_C$ , and  $R$ ) are added vectorially, the resultant will be the circuit *impedance*. By definition, impedance is the combined opposition to the flow of alternating current that is offered by the inductance, capacitance, and resistance in that circuit.

The qualities of inductive and capacitive reactance give rise to an important area of applications in what are referred to as tuned circuits. The name *tuned circuits* generally refers to a combination of inductive and capacitive reactances arranged so that the impedance of the combination changes abruptly at some specific frequency, or narrow range of frequencies. Tuned circuits are made with inductors and capacitors. As with any circuit, they also have resistance. Resistance is frequently useful in establishing the desired performance characteristics of a tuned circuit, and it is present in all circuits. The frequency at which the tuned circuit exhibits its sharp change in impedance is referred to as the *resonant frequency*. There is also a term *antiresonant frequency* which, in some applications, should be used instead of the term resonant frequency

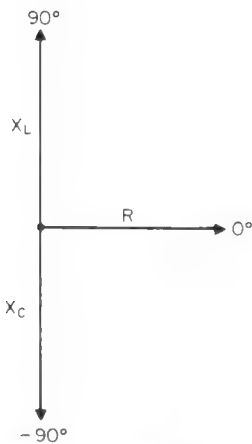


Fig. 10.1 Graphical representation of  $X_L$ ,  $X_C$ , and R.

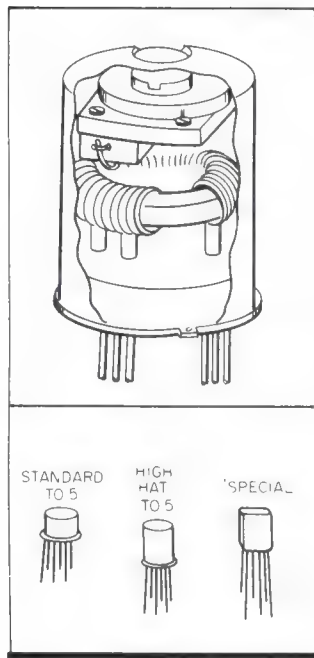


Fig. 10.2 A miniaturized tuned circuit, used for printed-circuit boards. In the upper view the microminiature, variable ceramic capacitor is at the top. Beneath this is the inductor, wound on a powdered iron toroidal core. (Courtesy JFD Electronics Corp.)

when referring to certain types of tuned circuits. There is a fine distinction between the terms which will be covered later in this chapter.

**ANALOGIES** There are many analogies between electronic tuned circuits and other physical phenomena. The relationship between the length and period of swing for a pendulum is one example. A tuning fork is another good example of a tuned mechanical device. A tuning fork will vibrate at only one frequency, regardless of how hard it is struck, or how it is set into vibration. In a like manner, an electrical tuned circuit will operate at only one frequency, or at only one narrow band of frequencies.

**SOME USES OF TUNED CIRCUITS** Tuned circuits are an extremely important factor in the overall art and science of electronics, as they make it possible to perform many functions that would otherwise be either impossible or extremely difficult. These functions include *selective amplification of certain frequencies* (or bands of frequencies) without amplifying other unwanted frequencies, and *rejection of unwanted frequencies* or bands of frequencies. Tuned circuits are utilized in radio broadcast receivers as the means of selecting the one desired station to be received from among the large number of stations that are simultaneously delivering signals to the receiver. Tuned circuits are employed in radio transmitters to establish the transmitter carrier frequency. In television receivers they are used to select the desired channels. In electronic circuits there are many other applications.

The importance of tuned circuits in the overall field of electronics can hardly be overestimated. It can safely be said that, were it not for tuned circuits, the art and science of electronics as we know it today would not exist. Figure 10.2 shows an

example of a tuned circuit so small that it can be packaged into a transistor case. Much larger tuned circuits are used in transmitters where much larger powers are involved.

## 10.2 TYPES OF TUNED CIRCUITS

There are many different methods of classifying tuned circuits. For example, they may be classified as *tunable* or *fixed*—depending on whether the frequency of the tuned circuit can be varied, or whether it is always set at the same frequency. They may be called *narrow-band* or *wideband*—depending on the shape of their characteristic curves. They are referred to as being either *series-tuned* or *parallel-tuned* circuits—according to whether their components are series- or parallel-connected. Further, a given tuned circuit can be classified according to more than one of the several different methods of distinction. That is, a tuned circuit may be a tunable, narrow-band, parallel-tuned circuit.

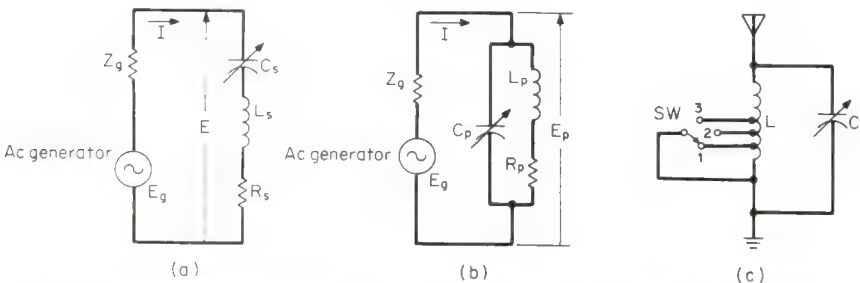
A few examples of commonly used tuned circuits are illustrated in Fig. 10.3. The most often used circuit consists of a variable capacitor either in series or in parallel with an inductor (Fig. 10.3a and b). The minimum and maximum values of capacitance for the variable capacitor are chosen to ensure complete coverage of a band of frequencies. Where a number of different frequency bands must be tuned, as in a short-wave receiver, a switchable inductor may be used (Fig. 10.3c). If, for example, the switch is in position one, the capacitor and resulting inductance covers the lower-band (broadcast) frequencies. In position two or three, less inductance is present across the capacitor; for these positions the circuit is tuned to the shortwave frequencies.

The distinction between *fixed* (sometimes called *fixed-tuned*) and *tunable* (sometimes called *variable*) circuits refers to whether the circuit is tuned permanently to a single (*fixed*) frequency, or whether the frequency at which the circuit exhibits its unique impedance-response characteristic can be altered, or varied, according to the needs and desires of the user.

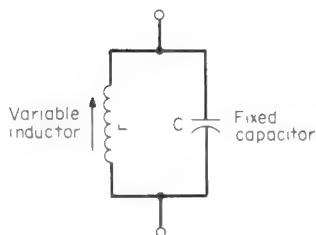
The many different methods of classifying tuned circuits, and the fact that a given tuned circuit can fit simultaneously into more than one classification, may seem confusing at first. However, there is a reason for each of the different methods of classification, and each has its own importance. This will become further apparent after the circuits have been discussed in greater detail.

**FIXED-TUNED CIRCUITS** Fixed-tuned circuits are set permanently to a single frequency. They find many applications in electronic equipment where it is desired to respond to a known, predetermined frequency, either by enhancing the equipment response to that particular frequency as compared to the response to other frequencies, or by producing an especially *low* response to (rejecting) the frequency to which the circuit is resonant.

Some nominally fixed-tuned circuits may, however, be slightly *adjustable*. The term *adjustable* generally refers to the fact that one or more of the reactive elements of the tuned circuit can be varied (adjusted) over a narrow range. This adjustment is most often provided for the purpose of compensating for minor variations or imperfections



**Fig. 10.3** Examples of tuned circuits: (a) A series-tuned circuit. (b) A parallel-tuned circuit. (c) Tuned circuit used in radio receiver circuit. Switch permits the receiver to be tuned to more than one band of frequency.



**Fig. 10.4** A tuned circuit in which the resonant frequency is adjusted with a variable inductor.

of the circuit elements (or in some cases, minor variations in frequency of the input signal that is desired to be tuned). The frequency of the tuned circuit in Fig. 10.4 is adjusted with a slug-tuned inductor.

**TUNABLE CIRCUITS** Tuned circuits which are slightly adjustable circuits should not be confused with tunable circuits, as the latter have a very different kind of application. In general, tunable, or variable, circuits are used in applications in which it is desirable to be able to change the frequency for which the circuit exhibits its characteristic *selective* response. Such changing may be done frequently and with relative ease. Perhaps the most common example of the use of tunable circuits is in radio broadcast receivers, in which tunable input circuits can be varied by the user to select the particular one station to be received out of the many that are broadcasting simultaneously on different frequencies.

**CIRCUIT BANDWIDTH** Another method of classifying tuned circuits is as “narrow-band” or “wideband”—whether the circuit responds only to frequencies within a very narrow range or *band* about the resonant frequency, or whether the response is to frequencies in a relatively wide band about the resonant frequency. (The terms *narrow band* and *wide band* are, of course, very general, and more sophisticated methods for describing the bandwidth of tuned circuits are needed in most practical situations.)

The fact that a tuned circuit is narrow-band or wideband may sometimes be a matter of necessity brought on by the practical limitations of the components with which the equipment designer must work. More often, however, it is a matter of deliberate design to fit the requirements of the application. In a radio broadcast receiver, for example, the tuned circuits that determine which of the many available stations is actually received are narrow-band circuits designed so that the one desired station can be received clearly, without interference such as might be produced by other stations broadcasting on frequencies near that of the desired station.

In many applications, on the other hand, including television broadcast receivers, radar receivers, and other equipment that must handle a wide band of signals, the tuned circuits are deliberately made so as to respond effectively to signals within a relatively wide band on both sides of the center frequency to which the tuned circuit is resonant.

**SINGLE-TUNED AND DOUBLE-TUNED CIRCUITS** The distinction between single-tuned and double-tuned (or more generally, multiple-tuned) circuits refers to certain details of tuned-circuit configuration and/or adjustment. Generally speaking, the distinction concerns whether the tuned circuit is “resonant” to only one frequency (single-tuned), or whether it is resonant *simultaneously* to two or more frequencies (double- or multiple-tuned).

Multiple tuning of circuits may be done for a variety of reasons, but most often it is a means of obtaining wideband response without sacrificing other important circuit performance characteristics. One of these other performance characteristics is *selectivity*, which is a measure of the ability of a tuned circuit to respond to a desired frequency or band of frequencies, while *rejecting* frequencies outside this band, even though close to it.



**SERIES-TUNED AND PARALLEL-TUNED CIRCUITS** Perhaps the most important single method of classifying tuned circuits is based on the circuit configuration. As we shall see in the next section, a tuned circuit with the components connected in series may have a completely different function from one with the components connected in parallel.

### 10.3 COMPARISON OF SERIES-TUNED AND PARALLEL-TUNED CIRCUITS

The terms “series circuit” and “parallel circuit” refer to the arrangement of the capacitive and inductive elements that make up the tuned circuit. The distinction can readily be understood with the aid of Fig. 10.3. Figure 10.3a shows a simple series-tuned circuit, and Fig. 10.3b shows a parallel-tuned circuit.

In Fig. 10.3a the capacitor  $C_s$  and the inductor  $L_s$  that make up the tuned circuit are in series across the source of input voltage (shown schematically as a generator in Fig. 10.3a). Also in series with the other circuit elements are the resistances of the tuned circuit, shown schematically as a single lumped resistance  $R_s$  and the internal impedance of the generator, or *source generator*,  $Z_g$ .

In Fig. 10.3b the capacitor  $C_p$  and the inductor  $L_p$  are in parallel across the “source” generator with its internal impedance  $Z_g$ . In this parallel-tuned circuit the symbol  $R_p$  is used to denote the *total* effective resistance of the circuit. This lumping of resistance into a single equivalent-circuit element is common practice, and reflects the practical fact that, in the great majority of instances, the resistance of the inductor is so much greater than that of the capacitor that the latter is, in fact, negligible.

Figure 10.3, then, illustrates the basic configurations of both series- and parallel-tuned circuits. The origin of the terminologies for describing these two types of tuned circuits is apparent from the illustration. Note that the distinguishing feature of the series circuit is that all the generator current *must* flow through  $R$ ,  $L$ , and  $C$ . In the parallel-tuned circuit, the generator current divides so that each branch receives only part.

### 10.4 SERIES-TUNED CIRCUITS

**APPLICATIONS** Series-tuned circuits have applications wherever it is desired to pass a signal of one frequency with minimum attenuation while rejecting signals of all other frequencies.

As one example, consider a radio transmitter that is to be coupled to its antenna in such a way that the desired transmitter output frequency is transmitted without attenuation, but the harmonics—that is, multiples—of the output frequency that may also be present do not reach the antenna. This can be accomplished by using a series-tuned circuit, resonant at the desired output frequency, between the transmitter output and the antenna, as shown in Fig. 10.5. The series-tuned circuit, being resonant at the desired carrier frequency, will have minimum effect on the transmitter output at this frequency. At higher frequencies, however, such as at the second harmonic (second multiple) of the desired output frequency, the series circuit will present an impedance approximately equal to the reactance of the inductor at the frequency in question, thus effectively preventing harmonics from reaching the antenna.

Another example of the use of series-tuned circuits may be found in what are frequently referred to as “trap” circuits. These are circuit applications in which a series-

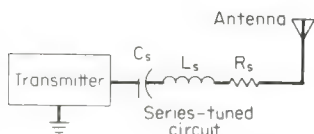


Fig. 10.5 Use of a series-tuned circuit in transmitter output.

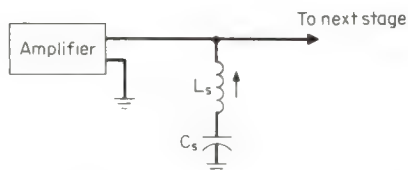
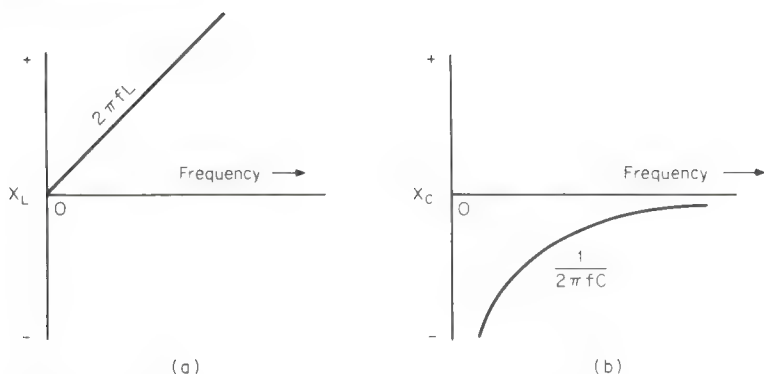


Fig. 10.6 An example of a series-tuned circuit used as a trap across the output of an amplifier stage.





**Fig. 10.7** Comparison of inductive and capacitive reactance curves: (a) Inductive reactance versus frequency. (b) Capacitive reactance versus frequency.

tuned circuit is used in parallel, or shunt, across a source of voltage, for the purpose of short-circuiting, or "trapping," signals of one frequency, while having negligible effect on signals of other frequencies. A specific example of the use of a series-tuned circuit as a trap circuit is shown in Fig. 10.6. This type of trap can be found in many television receivers.

The characteristic curves for inductors and capacitors are shown in Fig. 10.7.

**FUNDAMENTALS** The *reactance* of an *inductor* varies with frequency according to the equation

$$X_L = 2\pi fL \quad (10.1)$$

where  $X_L$  = inductive reactance,  $\Omega$

$f$  = applied frequency, Hz

$L$  = value of inductance, H

$\pi = 3.14$  (approximately)

Equation (10.1) states that for a given value of inductance  $L$  the inductive reactance  $X_L$  varies directly with frequency. This is shown graphically in Fig. 10.7a. At zero frequency (dc) the reactance is zero and the inductor acts as a short circuit. As the frequency increases, the inductive reactance increases in a direct manner.

The *reactance* of a *capacitor* varies with frequency according to the equation

$$X_C = -\frac{1}{2\pi fC} \quad (10.2)$$

where  $X_C$  = capacitive reactance,  $\Omega$

$C$  = value of capacitance, F

Equation (10.2) states that the capacitive reactance varies *inversely* with frequency; that is, as the frequency increases the capacitive reactance decreases—opposite to that for an inductor. Referring to Fig. 10.7b, at frequencies close to zero it can be seen that the capacitive reactance has a very large value, approaching infinity. The capacitor at zero frequency (dc), therefore, acts as an open circuit. At higher frequencies  $X_C$  approaches zero, and the capacitor is said to act as a short circuit.

The right side of Eq. (10.2) is preceded by a minus sign to denote the fact that capacitive reactance is opposite in effect to inductive reactance, and is considered to be *negative* reactance, whereas inductive reactance is considered to be *positive*. Equations (10.1) and (10.2) and also Fig. 10.7a and b both show that inductive reactance and capacitive reactance are oppositely affected by frequency. That is, *inductive reactance increases* as frequency increases, whereas the *value of capacitive reactance decreases* as frequency increases.

For a given combination of values of inductance  $L$  and capacitance  $C$  in a series circuit there will be some one frequency for which the inductive reactance  $X_L$  has the

same magnitude as the capacitive reactance  $X_C$ . If we remember that inductive reactance and capacitive reactance are opposites, that is, the former is positive while the latter is negative, it follows that they tend to cancel each other when they are equal in value.

The frequency for which the value of inductive reactance  $X_L$  equals the value of capacitive reactance  $X_C$  is the *resonant* frequency of a tuned circuit consisting of an inductance of  $L$  henrys and a capacitance of  $C$  farads. In other words, the resonant frequency of an  $LC$  series circuit is the frequency for which

$$X_L = X_C \quad (10.3)$$

The negative sign is omitted because, by definition, the resonant frequency is that frequency for which the inductive reactance is *numerically* equal to the capacitive reactance.

Substituting  $2\pi fL$  for  $X_L$ , and  $\frac{1}{2\pi fC}$  for  $X_C$ , we get

$$2\pi f_r L = \frac{1}{2\pi f_r C} \quad (10.4)$$

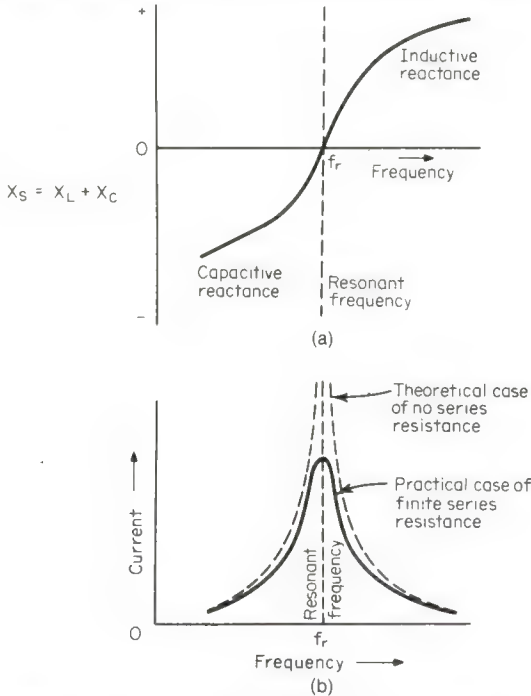
where  $f_r$  = resonant frequency, Hz.

Solving Eq. (10.4) for  $f_r$ , we obtain

$$f_r = \frac{1}{2\pi\sqrt{LC}} \quad (10.5)$$

Equation (10.5) states that the resonant frequency  $f_r$  is inversely proportional to the square root of the  $LC$  product. Thus, for a high resonant frequency the  $LC$  product is small, and for a low resonant frequency it is large.

The behavior of the series-resonant circuit is summarized in Fig. 10.8. Figure 10.8a



**Fig. 10.8** Characteristic curves for a series-resonant circuit: (a) Reactance of a series-tuned circuit versus frequency. (b) Current through a series-tuned circuit at frequencies in the vicinity of resonance.

## 10-8 Tuned Circuits

shows that at the resonant frequency the net reactance is zero, that is,  $X_L = X_C$ . Below the resonant frequency, capacitive reactance predominates; at frequencies above resonance, inductive reactance is predominant.

From Fig. 10.8b it is seen that the current in a series-tuned circuit is maximum at the resonant frequency. For frequencies less than or greater than the resonant frequency, the current falls and has a lower value. The magnitude of the current at resonance depends on the series resistance present. If there is zero resistance, the current approaches infinity; as the series resistance increases, the current at resonance decreases.

**example 10-1** A series-tuned circuit, like the one shown in Fig. 10.3a, has an inductance of 3 mH, a capacitance of 30  $\mu$ F, and a resistance of 300  $\Omega$ . What is the resonant frequency of this tuned circuit?

**solution** In a series-tuned circuit the amount of resistance in the circuit can be disregarded. (This is not always true in a parallel-tuned circuit.) To solve the problem, the inductance and capacitance values are expressed in henrys and farads.

$$3 \text{ mH} = 3 \times 10^{-3} \text{ H}$$

$$30 \text{ } \mu\text{F} = 3 \times 10^{-5} \text{ F}$$

Substituting these values into Eq. (10.5), we get

$$\begin{aligned} f_r &= \frac{1}{2\pi\sqrt{LC}} \\ &= \frac{1}{(2 \times 3.14) \sqrt{3 \times 10^{-3} \times 3 \times 10^{-5}}} \\ &= \frac{1}{6.28 \sqrt{9 \times 10^{-8}}} \\ &= \frac{1}{6.28 \times 3 \times 10^{-4}} = 530 \text{ Hz} \end{aligned} \quad (10.5)$$

**example 10.2** It is desired to design a series-tuned circuit to have a resonant frequency of 5000 Hz. If the capacitor to be used has a value of 0.1  $\mu$ F, what value of inductance is required?

**solution** Substitution of the given values in Eq. (10.4) yields

$$2\pi f_r L = \frac{1}{2\pi f_r C} \quad (10.4)$$

or

$$2\pi(5000)L = \frac{1}{2\pi(5000) \times 0.1 \times 10^{-6}}$$

Solving for  $L$ ,

$$\begin{aligned} L &= \frac{1}{4\pi^2 \times 25 \times 10^6 \times 0.1 \times 10^{-6}} \\ &= \frac{1}{10\pi^2} = 0.01014 \text{ H} = 10.14 \text{ mH} \end{aligned}$$

The simple tuned circuit in Example 10.1 will be resonant at a frequency of 530 Hz. At this frequency, the inductive and capacitive reactances *cancel*. This means that the only opposition to current flow will be what is offered by  $Z_p$  and  $R_s$  (see Fig. 10.3a). If these values are small, then there will be very little opposition to current flow, and the circuit current will be very high.

Equation (10.5) applies to series-tuned circuits with or without a series resistance  $R_s$ . However, the equation will only apply to parallel circuits if the resistances in the  $L_p$  and  $C_p$  branches are small enough to be neglected. If, for example, the resistance value of  $R_p$  in Fig. 10.3b is more than about 10 percent of the value of  $X_L$ , then Eq. (10.5) *cannot be used* to determine the value of resonant frequency for the circuit. The same is true if a resistance is placed in series with  $C_p$  in the circuit.

**SERIES-TUNED VOLTAGES AT RESONANCE** Another property of series-tuned circuits that deserves attention concerns the voltages that exist across the circuit elements at frequencies at or near the resonant frequency of the circuit.

If we remember that, at the resonant frequency, the capacitive reactances  $X_C$  of the inductor are equal and opposite to each other; it follows then that the net reactance of

the series circuit will be zero at the resonant frequency. The current flowing through the circuit at the resonant frequency is thus given by the simple Ohm's law relationship,

$$I = \frac{E}{R_s} \quad (10.6)$$

where  $E$  = voltage applied across the entire series-tuned circuit, that is, the voltage developed by the voltage source less the voltage drop across the internal impedance of the source ( $Z_g$  in Fig. 10.3a)

$I$  = current through the series-tuned circuit at resonance

$R_s$  = total series resistance of the series-tuned circuit

If the total series resistance  $R_s$  is reasonably low, as it will be in a well-designed series-tuned circuit for most applications, the current  $I$  through the circuit will be quite large at resonance.

The voltage across the capacitor at resonance is given by the Ohm's law relationship,

$$E_C = I \times X_C \quad (10.7)$$

where  $E_C$  = voltage across the capacitor at resonance.

Similarly, the voltage across the inductor at resonance will be

$$E_L = I \times X_L \quad (10.8)$$

where  $E_L$  = voltage across the inductor at resonance.

The last two equations state that for a series-tuned circuit at resonance, the voltage across the capacitor (or inductor) is equal to the product of current  $I$  at resonance [defined by Eq. (10.6)] and the capacitive reactance (or inductive reactance). The practical significance of these effects will be considered shortly. Equations (10.7) and (10.8) both assume that whatever resistance is present in the series-tuned circuit is considered as a separate element. However, this assumption does not in any way affect the validity of the above equations.

Bearing in mind that, in a series-tuned circuit with a typically low value of series resistance  $R_s$ , the current through the circuit at resonance is quite large [see Eq. (10.6)], it follows that the voltages across the capacitive and inductive elements of the circuit are also quite large at resonance. These voltages, which will be equal in magnitude, will be greater than the voltage applied across the series-tuned circuit by a factor of  $X_C/R_s$ , or  $X_L/R_s$ . That is, if the voltage applied across the total circuit at resonance is  $E$ , the voltage across the capacitor and inductor may be also expressed as

$$E_C = E \left( \frac{X_C}{R_s} \right) \quad (10.9)$$

$$\text{and} \quad E_L = E \left( \frac{X_L}{R_s} \right) \quad (10.10)$$

The high voltages across the reactive components affect the voltage ratings. They must be considered if the circuit is to operate without arcing or short circuiting. The ratio of capacitive or inductive reactance at resonance to the equivalent series resistance may have values in the range of from 10 to perhaps 100; so it follows that the capacitor and inductor must be constructed to withstand voltages from 10 to 100 times as great as the voltage applied across the series-tuned circuit. (This refers, of course, to the r-f or ac voltage applied to the circuit, and not to any dc voltage that may be present across the series-tuned circuit.)

In low-voltage applications, such as a series-tuned circuit in a radio receiver, this is not a matter of great consequence. In other applications, such as in transmitters where the applied voltage itself may be reasonably large, the requirement that the individual elements of the series-tuned circuit withstand voltages that are several times the applied voltage can represent an important design requirement. This is illustrated in the next example.

**example 10.3** In a series-tuned circuit  $R_s = 10 \, \Omega$ , and the magnitude of the capacitive reactance is  $1000 \, \Omega$  at the resonant frequency. If the voltage source  $E_g = E = 10 \, \text{V}$ ,  $Z_g \approx 0$ , what are the voltages across the inductor, capacitor, and resistor at the resonant frequency?

## 10-10 Tuned Circuits

**solution** From Eq. (10.6),

$$I = \frac{E}{R_s} = \frac{10}{10} = 1 \text{ A}$$

At resonance,

$$X_C = X_L = 1000 \Omega$$

Using Eqs. (10.7) and (10.8), we find

$$\begin{aligned} E_C &= I \times X_C \\ &= 1 \times 1000 \\ &= 1000 \text{ V} \end{aligned} \quad \text{and} \quad \begin{aligned} E_L &= I \times X_L \\ &= 1 \times 1000 \\ &= 1000 \text{ V} \end{aligned}$$

Also,

$$E = I \times R_s = 1 \times 10 = 10 \text{ V}$$

The significance of this example is that, with practical values of circuit elements, the individual voltages across the capacitor and inductor at resonance are *100 times* the voltage applied across the series-tuned circuit.

## 10.5 PARALLEL-TUNED CIRCUITS

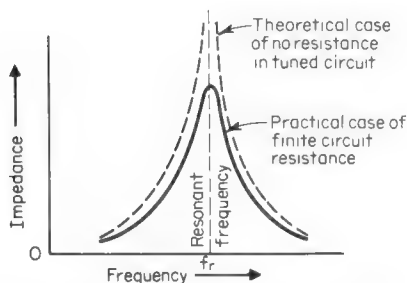
In the preceding section, the characteristics and behavior of series-tuned circuits were described and discussed. We now consider the second major category of tuned circuits, the parallel-tuned circuit, which is also sometimes referred to as an *antiresonant circuit*.

In Fig. 10.3b, the parallel-tuned circuit consists essentially of a capacitor  $C_p$  and an inductor  $L_p$  connected in parallel (or *shunt*) across a source of voltage, a *generator*. The circuit in Fig. 10.3b shows resistance in the inductive branch, but not in the capacitive branch. This is a practical consideration because the resistance of the wires in the inductor is usually large compared to the ESR (Equivalent Series Resistance) of the capacitor. However, this is not to say that resistance will never be found in the capacitive branch. The equation for parallel resonance will be developed later in this chapter for all possible cases.

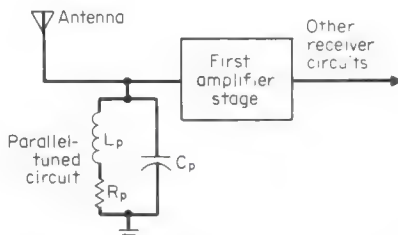
**APPLICATIONS** Applications of parallel-tuned circuits are based on their impedance versus frequency characteristics, as illustrated in Fig. 10.9. At the resonant frequency the impedance is maximum, and at frequencies below and above resonance the impedance falls off. In many applications parallel-tuned circuits are used to obtain selective filtering of specific frequencies.

Figure 10.10 illustrates the use of a parallel-tuned circuit to obtain selective amplification of a desired frequency in a radio or television receiver. The antenna is connected to the parallel-tuned circuit, to which the input terminal of the first amplifier stage is also connected.

When signals arrive from the antenna at the resonant frequency of the parallel-tuned circuit, this presents a high impedance to ground. Thus, the signal current supplied



**Fig. 10.9** Impedance of a parallel-tuned circuit at frequencies in the vicinity of resonance.



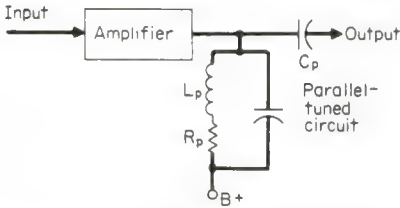
**Fig. 10.10** Use of a parallel-tuned circuit for selective amplification at the receiver input.



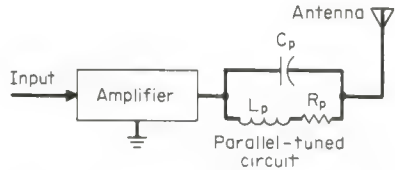
by the antenna develops a relatively high voltage across the tuned circuit, this voltage being amplified by the first amplifier stage.

For signals of other frequencies, that is, frequencies relatively far from the resonant frequency, the impedance is much lower. The voltage developed in response to the current from the antenna, therefore, will also be lower, and the signals will not be amplified to any appreciable degree. (In fact, the circuit of Fig. 10.10 may be viewed as one that has the effect of *shunting* all signals (other than those at the resonant frequency) to ground. Since they are shunted to ground, they are not presented to the amplifier.

A similar application of parallel-tuned circuits is illustrated in Fig. 10.11. A parallel-tuned circuit is used at the output of an amplifier, instead of at the input as in Fig. 10.10.



**Fig. 10.11** Use of a parallel-tuned circuit for selective amplification at the amplifier output.



**Fig. 10.12** Use of a parallel-tuned circuit in series with amplifier output.

The amplifier in Fig. 10.11 may be a circuit in a transmitter, or in a receiver. In either case, the purpose of the circuit is to cause the amplifier to respond primarily to the resonant frequency of the tuned circuit, and to minimize the amplification of signals at other frequencies.

The voltage gain of an amplifier is directly proportional to the impedance of the load. In the circuit of Fig. 10.11, the parallel-tuned circuit is the amplifier load. At the resonant frequency this impedance will be quite high, and therefore the gain will be high. At all other frequencies the impedance of the parallel-tuned circuit will be low, and as a result the gain will be low. To summarize, the parallel-tuned circuit permits the amplifier to produce a large voltage gain at a specific frequency.

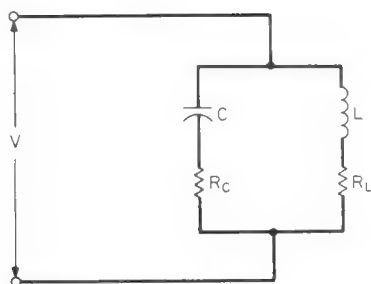
Another type of application of parallel-tuned circuits is illustrated in Fig. 10.12. In this circuit the parallel-tuned circuit is in series with an amplifier load (which in this case is an antenna). Its purpose is to prevent signals of an undesired frequency from reaching the load or the next stage. This is accomplished because the parallel-tuned circuit presents a high impedance to the signals of the (undesired) frequency to which it is tuned, while presenting a relatively low impedance to signals of the desired frequency. The circuit of Fig. 10.12 is used frequently in radio transmitters to prevent the second harmonic (second multiple) of the desired output frequency from reaching the antenna, from which it could be radiated and cause interference.

**FUNDAMENTALS** The equation for parallel resonance (which is sometimes called *antiresonance*) will be obtained for the circuit of Fig. 10.13. A few mathematical derivations are given in this section, but the overall results are summarized for those who want a less mathematical approach to the subject of parallel-tuned circuits. Note that the resistances in both the inductive and capacitive branches will be taken into consideration. Then the equation will be simplified for use in certain specialized cases such as the one shown in Fig. 10.3b.

When a circuit is in resonance, the current and the voltage in the circuit must be in phase. This is true for both series and parallel resonant circuits. Another way of stating the condition for resonance is that the circuit must have *unity power factor*.

The only way that unity power factor can occur in the circuit of Fig. 10.13 is if the phasor sum of the currents in the two branches produces a total current that is in phase with the applied voltage  $V$ . The mathematics required to establish the values of  $R_L$ ,  $C$ ,  $R_C$ , and  $L$  that are needed to produce resonance would be very tedious. Fortunately, there is a shortcut method.





**Fig. 10.13** A general circuit for determining parallel resonance.

The two series branches of Fig. 10.13 can be converted to the four parallel branches shown in Fig. 10.14. The equations required for finding the conductances  $G_C$  and  $G_L$  and the susceptances  $B_C$  and  $B_L$  are shown in the illustration. It is important to understand that the conversion from the circuit of Fig. 10.13 to the one in Fig. 10.14 is a standard procedure used for simplifying and solving parallel circuits. In other words, this is not a special case, but rather it is a standard procedure.

In the equivalent circuit of Fig. 10.14 there is no resistance in the capacitive and inductive branches. Resonance will occur when  $B_C = B_L$ , and this is the basis for finding the equation for parallel resonance. (Compare this with the solution of the series-resonant equation which is based on  $X_C = X_L$ .)

$$B_L = B_C$$

$$\frac{X_L}{R_L^2 + X_L^2} = \frac{X_C}{R_C^2 + X_C^2}$$

$$\frac{2\pi fL}{R_L^2 + (2\pi fL)^2} = \frac{1/2\pi fC}{R_C^2 + (1/2\pi fC)^2}$$

When this equation is solved for  $f_r$ , which is the resonant frequency of the circuit in Fig. 10.13, the result is

$$f_r = \frac{1}{2\pi\sqrt{LC}} \sqrt{\frac{L - CR_C^2}{L - CR_L^2}} \quad (10.11)$$

where  $f_r$  = resonant frequency, Hz

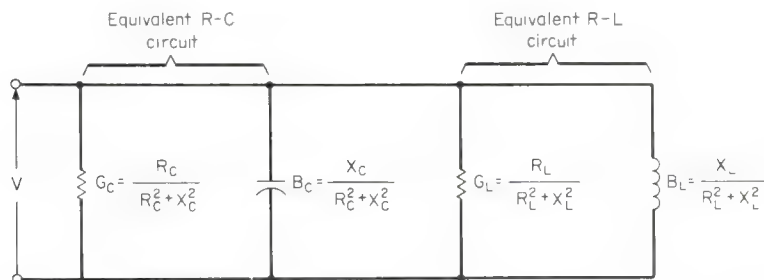
$L$  = circuit inductance, H

$C$  = circuit capacitance, F

$R_C$  = resistance in capacitive branch,  $\Omega$

$R_L$  = resistance in inductive branch,  $\Omega$

This is the equation for the resonant frequency of the circuit in Fig. 10.13, even though it is derived from the equivalent circuit of Fig. 10.14.



**Fig. 10.14** This circuit has the same impedance and will produce the same phase angle as the circuit in Fig. 10.13.

Several very important facts about parallel resonance can be determined from Eq. (10.11).

1. When the resistance in the two branches is negligible, the value zero can be substituted for both  $R_c$  and  $R_L$  in the equation. The result is

$$f_r = \frac{1}{2\pi\sqrt{LC}} \quad (10.12)$$

where  $f_r$  = resonant frequency, Hz

$L$  = inductance, H

$C$  = capacitance, F

In other words, the same equation can be used for both series and parallel resonance if there is negligible resistance in the parallel branches. In actual practice, if  $X_L$  is at least ten times the value of  $R_L$ , and  $X_C$  is at least ten times the value of  $R_c$ , the resistances are considered to be negligible.

2. Equation (10.11) shows that a parallel-resonant circuit can actually be tuned by varying the resistance in one of the branches—that is, by varying either  $R_c$  or  $R_L$ . This follows from the fact that changing the value of either  $R_c$  or  $R_L$  in the equation will change the value of  $f_r$ . (Of course,  $f_r$  can also be varied by changing either  $L$  or  $C$ .)

If we solve Eq. (10.11) for the value of  $R_L$ , the result is

$$R_L = \sqrt{\frac{X_L}{X_C} R_c^2 - X_L^2 + \frac{L}{C}} \quad (10.13)$$

where the letters have the same meanings as in Eq. (10.11).

If we solve Eq. (10.11) for the value of  $R_c$ , the result is

$$R_c = \sqrt{\frac{R_L^2}{(2\pi f)^2 LC} - \frac{1}{\omega^2 C^2} + \frac{L}{C}} \quad (10.14)$$

where the letters have the same meanings as in Eq. (10.11).

Equations (10.13) and (10.14) give the value of  $R_L$  or the value of  $R_c$  needed to bring the circuit of Fig. 10.13 into resonance.

3. It was mentioned earlier that the circuit of Fig. 10.3b is often given as a practical circuit because the dc resistance of the inductive branch cannot always be ignored. The equation for resonance in this circuit can be obtained by setting  $R_c = 0$  in Eq. (10.11). The result is

$$f_r = \frac{1}{2\pi\sqrt{LC}} \sqrt{\frac{L - CR_L^2}{L}} \quad (10.15)$$

4. In a series  $LC$  circuit there is always a resonant frequency because there is always some frequency at which  $X_L = X_C$ . However, in the parallel-resonant circuit, it is possible to have a combination of components that are not resonant at any frequency. To understand how this is possible, look at the numerator and denominator under the radical in Eq. (10.11):

Numerator:	$L - CR_L^2$
Denominator:	$L - CR_c^2$

A very important point to consider is that the value under the radical will be negative if

	$CR_L^2$ is greater than $L$
or	$CR_c^2$ is greater than $L$

If the value under the radical is negative, no value of  $f_r$  exists because the result is an *imaginary* value. To summarize, then, it is possible to have a parallel circuit that has no resonant frequency!

5. Another important feature of parallel-resonant circuits is that they may be designed to be resonant at all frequencies. This can be demonstrated with a short mathematical derivation.

In the derivation of Eq. (10.11) this equation was given:

$$\frac{2\pi fL}{R_L^2 + (2\pi fL)^2} = \frac{1/2\pi fC}{R_c^2 + (1/2\pi fC)^2}$$

10-14 Tuned Circuits

To simplify equations of this type, the symbol  $\omega$  (omega) is substituted for  $2\pi f$

$$\frac{\omega L}{R_L^2 + \omega^2 L^2} = \frac{1/\omega C}{R_C^2 + 1/\omega^2 C^2}$$

Setting the product of the means equal to the product of the extremes (sometimes called "cross-multiplying") gives

$$\omega L R_C^2 + \frac{\omega L}{\omega^2 C^2} = \frac{R_L^2}{\omega C} + \frac{\omega L^2}{C}$$

The lowest common denominator is  $\omega C^2$ . Multiplying every term by this value gives

$$\omega^2 L C^2 R_C^2 + L = R_L^2 C + \omega^2 L^2 C$$

Now, if either the numerator or the denominator under the radical in Eq. (10.11) is equal to zero, then the resonant frequency cannot be given a numerical value.

For the numerator:

$$L - C R_L^2 = 0$$
$$R_L = \sqrt{\frac{L}{C}}$$

For the denominator:

$$L - C R_C^2 = 0$$
$$R_C = \sqrt{\frac{L}{C}}$$

Substituting the radical equivalent values for  $R_L$  and  $R_C$  in the above equations gives

$$\frac{\omega L^2}{C} + \frac{L}{\omega C^2} = \frac{\omega L^2}{C} + \frac{L}{\omega C^2}$$

Note that this equation is valid for *every value* of  $\omega$  (and therefore every value of frequency  $f$ )—*except the value of  $\omega = 0$* . The importance of this derivation is that the circuit is resonant at *all frequencies* whenever  $R_L = R_C = \sqrt{L/C}$ .

In a series-tuned circuit, if the applied frequency is above resonance, there is a greater voltage drop across the inductor than across the capacitor. This makes the circuit behave like an inductor. If the applied frequency is below the resonant frequency, on the other hand, the circuit behaves like a capacitor because the larger voltage drop is across the capacitor.

TABLE 10.1 Comparison of Series- and Parallel-Tuned Circuits

	Series-tuned circuit	Parallel-tuned circuit
Impedance	Minimum at the resonant frequency	Maximum at the resonant frequency
Current	Maximum at resonant frequency	Minimum at resonant frequency
Voltage	Minimum voltage across the circuit at resonant frequency	Maximum voltage across the circuit at resonant frequency
Resonance	There is one resonant frequency for each LC combination. Resistance does not affect resonance for all practical purposes.	a. Resistance affects resonant frequency. b. Circuit may resonate at all frequencies. c. Circuit may not have a resonant frequency.
When applied frequency is below resonance	Circuit is capacitive.	Circuit is inductive.
When applied frequency is above resonance	Circuit is inductive.	Circuit is capacitive.
Power factor at resonance	1.0	1.0
Special feature	Voltage across a component may be many times applied voltage.	Current through a component may be many times main current.

When the frequency of the applied signal across a parallel-tuned circuit is above the resonant frequency, more current flows through the capacitor because its reactance is lower than that of the inductor. This means that the circuit will be capacitive. Conversely, if the applied frequency is below resonance, the greater current flows through the inductor, and the circuit is inductive.

The characteristics of series- and parallel-tuned circuits are summarized in Table 10.1.

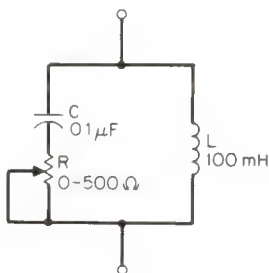


Fig. 10.15 Circuit for Example 10.5.

**example 10.4** In the parallel-tuned circuit of Fig. 10.13 assume that  $L_p = 100$  mH,  $C_p = 0.1$   $\mu$ F, and  $R_C = R_L = 0$ . What is the resonant frequency of the circuit?

**solution** From Eq. (10.12),

$$\begin{aligned} f_r &= \frac{1}{2\pi\sqrt{LC}} \\ &= \frac{1}{6.28\sqrt{0.1 \times 0.1 \times 10^{-6}}} = 1590 \text{ Hz} \end{aligned} \quad (10.12)$$

**example 10.5** In the circuit of Fig. 10.15, a variable resistor is placed in the capacitive branch. This resistor is used for varying the resonant frequency. What is the range of frequencies through which  $R_C$  can tune the circuit?

**solution** When  $R_C = 0$ ,

$$f_r = \frac{1}{2\pi\sqrt{LC}} = 1590 \text{ Hz} \quad (\text{See Example 10.4.})$$

When  $R_C = 0.5$  k $\Omega$ ,

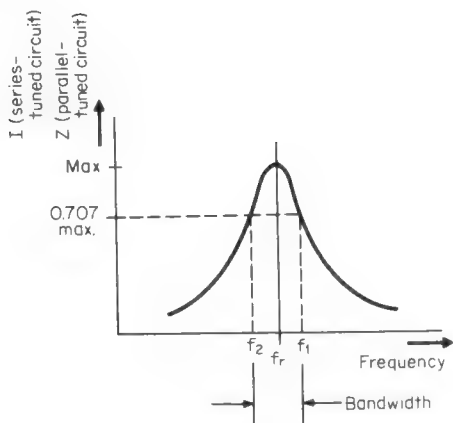
$$\begin{aligned} f_r &= \frac{1}{2\pi\sqrt{LC}} \frac{L}{L - CR_C^2} \\ &= 1590 \sqrt{\frac{100 \times 10^{-3}}{100 \times 10^{-3} - (0.1 \times 10^{-6})(0.5 \times 10^3)^2}} = 1828 \text{ Hz} \end{aligned}$$

Thus a change of 500  $\Omega$  results in a frequency change of 238 Hz.

## 10.6 SELECTIVITY, BANDWIDTH, AND Q FACTOR

The ability of a tuned circuit to respond differently to signals of different frequencies is referred to as its *selectivity*. For quantitative purposes, the selectivity of a tuned circuit—that is, its ability to discriminate between signals of different frequencies—is measured in terms of a parameter called *bandwidth*.

**BANDWIDTH** The bandwidth of a tuned circuit is a range of frequencies between two points on its selectivity curve—one on each side of the resonant frequency—at which the circuit response is a stated fraction of the response at the resonant frequency. The curve of Fig. 10.16 is an example of a selectivity curve. Although the fractional response point on the selectivity curve at which the tuned circuit bandwidth is determined can be stated arbitrarily, it is almost universal practice to measure bandwidth at points on the selectivity curve at which the circuit power response is one-half the



**Fig. 10.16** The bandwidth is taken between the points where the current, voltage, or impedance is 70.7 percent of maximum, depending on whether it is a series- or parallel-tuned circuit. It may also be taken between points where the power is 50 percent of maximum.

power that is produced at resonance. The bandwidth so measured is called the *half-power* bandwidth, or the *-3dB* bandwidth. The latter term is derived from the fact that one-half power corresponds to a power level three decibels below maximum power.

Although bandwidth is defined in terms of power, it is generally more convenient to measure it in terms of other circuit performance characteristics. The following relationships will be found especially useful in practice:

1. In a series-tuned circuit, the half-power bandwidth is the width (in frequency) of the portion of the selectivity curve between points at which the current through the circuit is equal to  $1/\sqrt{2} = 0.707$  times the current at resonance. (This relationship should be self-evident when it is remembered that power  $P$  is proportional to the square of current,  $P = I^2 R$ .)

2. In a parallel-tuned circuit, the half-power bandwidth is the width of the portion of the selectivity curve between points at which the effective parallel impedance of the circuit is equal to  $1/\sqrt{2} = 0.707$  times the impedance at resonance.

It is evident that the bandwidth relationships for series-tuned and parallel-tuned circuits correspond, with the difference that bandwidth of a series-tuned circuit is expressed in terms of current flow while bandwidth of a parallel-tuned circuit is expressed in terms of effective parallel impedance. Figure 10-16 summarizes the definitions of bandwidth for series- and parallel-tuned circuits.

**Q FACTOR.** The bandwidth of a tuned circuit is determined by a circuit parameter called  $Q$  or  $Q$  factor. Circuit  $Q$  is a very important parameter used for describing the performance characteristics of a tuned circuit.

The  $Q$  of a tuned circuit is equal to the reactance of either the capacitor or the inductor at the resonant frequency of the circuit; the reactance of the capacitor and inductor are, as explained previously, equal at the resonant frequency, divided by the effective series resistance of the circuit. This relationship applies to both the series- and parallel-tuned circuits of Fig. 10.3. For the series-tuned circuit,

$$Q = \frac{2\pi f_r L_s}{R_s} \quad (10.16)$$

Similarly, for the parallel-tuned circuit with negligible resistance,

$$Q = \frac{2\pi f_r L_p}{R_p} \quad (10.17)$$

Expressions similar to Eqs. (10.16) and (10.17) can be written in terms of the capacitive reactances at resonance in place of the inductive reactances. However, expressing  $Q$  in terms of the inductive reactance is usually more convenient.

Another way of expressing  $Q$  is in terms of the half-power bandwidth of the tuned circuit as related to its resonant frequency:

$$Q = \frac{f_r}{f_1 - f_2} = \frac{f_r}{\Delta f} \quad (10.18)$$

where  $f_1$  = frequency above the resonant frequency at which the half-power relationship exists (see Fig. 10.16)

$f_2$  = frequency below the resonant frequency at which the half-power relationship exists

$f_r$  = resonant frequency of a tuned circuit (either series or parallel)

$\Delta f = f_1 - f_2$  = bandwidth of the circuit

Equation (10.18) is especially convenient for determining the  $Q$  of a tuned circuit experimentally. It is necessary only to measure the frequencies at which the circuit exhibits a half-power response from which information, together with knowledge of the resonant frequency, the  $Q$  is readily calculated. (The required frequency measurement must, however, be made with some care for circuits with high  $Q$ 's because errors in measuring either  $f_1$  or  $f_2$  can lead to serious errors in determining bandwidth  $\Delta f$ , and therefore in calculating the  $Q$  of the circuit.)

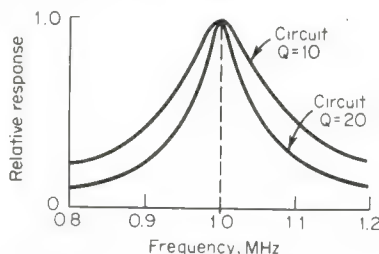


Fig. 10.17 Effect of circuit  $Q$  on bandwidth.

Measurements based on Eq. (10.18) are frequently superior to a method of determining  $Q$  according to Eq. (10.16) or (10.17). This is true because of the practical difficulty in determining the true value of  $R_s$  or  $R_p$ , inasmuch as their values are influenced by high-frequency losses in a manner that is not always easy to determine.

For use at many frequencies, however, there are instruments (usually called  $Q$  meters) that can measure the  $Q$  of an inductor, a capacitor, or a tuned circuit directly. This avoids the difficulty of determining the effective resistance by other means, and also the problems that sometimes arise in attempting to determine  $Q$  by frequency measurements.

It is evident from Eq. (10.16) or (10.17) that the resistance in a tuned circuit in relation to the reactance at resonance is the parameter that determines the circuit  $Q$ , and therefore the bandwidth. The effect of circuit  $Q$  on bandwidth is illustrated in Fig. 10.17. The higher the  $Q$ , the smaller the bandwidth.

The current in an inductor or a capacitor in a parallel-tuned circuit at resonance is  $Q$  times the external circuit current. Likewise, for a series-tuned circuit, the voltage across the inductor or capacitor at resonance is  $Q$  times the applied circuit voltage.

**example 10.6** In a parallel-tuned circuit,  $L = 100 \mu\text{H}$ ,  $R_L = 10 \Omega$ ,  $f_r = 1 \text{ MHz}$ . Determine the values of (a)  $Q$ , and (b) bandwidth.

**solution** (a) From Eq. (10.17),

$$Q = \frac{2\pi f_r L_p}{R_p} = \frac{6.28 \times 10^6 \times 100 \times 10^{-6}}{10} = 62.8 \quad (10.17)$$



## 10-18 Tuned Circuits

(There are no units for measuring  $Q$ .)

(b) Solving Eq. (10.18) for the bandwidth  $\Delta f$ ,

$$\Delta f = \frac{f_r}{Q} \quad (10.19)$$

Therefore

$$\Delta f = \frac{10^6}{62.8} \approx 16 \text{ kHz}$$

**example 10.7** Design a parallel-tuned circuit to cover the a-m broadcast band from 500 to 1600 kHz. At 500 kHz the minimum bandwidth should be 25 kHz. The minimum capacitance of the variable tuning capacitor is 30 pF, and stray wiring capacitance in parallel with the capacitor is on the order of 20 pF. Determine the value of  $L$ ,  $Q$ , and the capacitance range of the variable capacitor.

**solution** The value of  $L$  is determined at the upper frequency of 1600 kHz and at minimum capacitance of the capacitor, including stray capacitance ( $30 + 20 = 50$  pF). From Eq. (10.12),

$$L = \frac{1}{4\pi^2 f_r^2 C}$$

Substituting known values,

$$\frac{1}{4\pi^2 (1600 \times 10^3)^2 \times 50 \times 10^{-12}} \approx 200 \mu\text{H}$$

From Eq. (10.18),

$$Q = \frac{f_r}{\Delta f} = \frac{500}{25} = 20$$

By using the value of  $L = 200 \mu\text{H}$ , the maximum value of  $C$  is determined at the lower frequency of 500 kHz. Solving Eq. (10.12) for  $C$  gives

$$C = \frac{1}{4\pi^2 f_r^2 L}$$

and substituting known values,

$$= \frac{1}{4\pi^2 (500 \times 10^3)^2 \times 200 \times 10^{-6}} \approx 500 \text{ pF}$$

But 20 pF of stray capacitance is already present; therefore, the maximum value of capacitance is  $500 - 20 = 480$  pF, and the capacitor range is 30 to 480 pF.

**APPLICATION REQUIREMENTS** In some applications for tuned circuits the highest possible degree of selectivity—that is, the lowest bandwidth and the highest  $Q$ —is desired. This permits the circuit to do the best job of rejecting signals of unwanted frequencies, even though they may be very close to the desired frequency to which the circuit is tuned. In such cases, the tuned circuit is designed and constructed to preserve the highest possible  $Q$  by keeping the effective circuit resistance as low as possible.

There are also cases, namely in so-called wideband applications (of which television receiver circuits are an example), in which care must be taken lest the bandwidth of the tuned circuit be too narrow. This would result in some signals of desired frequencies being rejected, along with signals of undesired frequencies that may be farther from the resonant, or center, frequency of the circuit.

One method of keeping a tuned circuit from having a bandwidth that is too narrow is to deliberately introduce resistance into the circuit, thereby lowering the  $Q$  and producing a selectivity curve similar to the wider of the two curves in Fig. 10.17.

# Chapter 11

## Filters

### 11.1 INTRODUCTION

The term *filter* is used to describe a wide variety of circuits that are frequency-selective. Certain frequencies will pass through a given filter while others will be attenuated. Electric filters are analogous to lint filters on automatic dryers that allow air to pass through but trap the lint. Another analogy can be made with the fuel-line filter of an automobile. It traps dirt, but allows gasoline to go to the carburetor. By way of comparison, electric filters will pass some frequencies and trap (or attenuate) others.

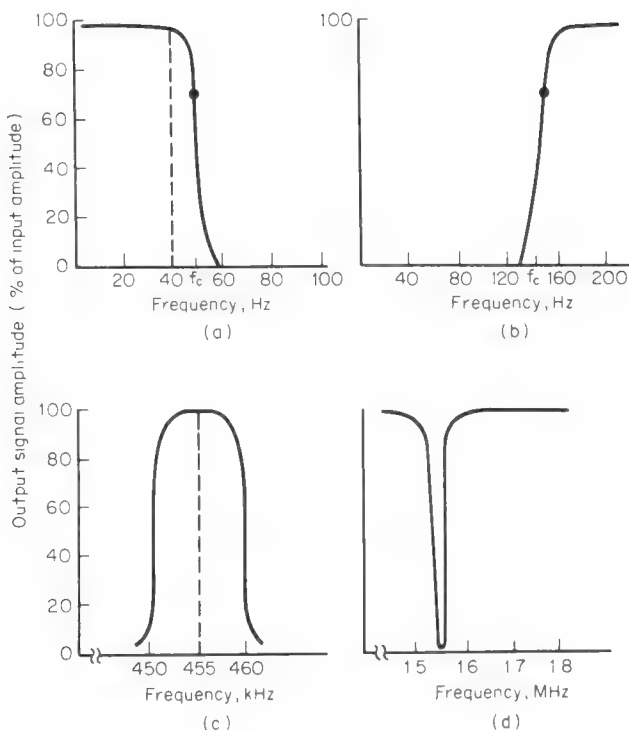
Knowing which frequencies are to pass and which are to be rejected, the circuit designer must decide which filter *configuration* is to be employed, and he must calculate the values of the components used. (The configuration is the method of electrically interconnecting the parts.) In this chapter we will discuss the advantages and disadvantages of the different configurations, and the methods used for determining their component values.

### 11.2 CLASSIFICATION OF FILTERS

There are many different ways of classifying or identifying filters. For example, they are sometimes identified by the layout of the components. Thus, a *pi filter* looks like the Greek letter  $\pi$ , and a *ladder circuit* looks like a ladder. Filters may be identified by the method used to design them. A *constant-k* filter is designed on the basis of a constant  $k$  which is related to the impedances of the circuit. An *m-derived filter* is based on a constant  $m$  which is derived from the constant  $k$ . Another way of identifying filters is by the shape of their characteristic curves. Thus, a *sharp cutoff filter* makes a very sharp distinction between which frequencies will pass through it and which will not. In some cases, the name of the person who originated the design or method of calculation is used for identifying the filter circuit. *Chebyshev filters* and *Butterworth filters* are examples of these.

One widely used scheme for classifying filters depends on the range of frequencies that the filter passes and rejects. According to this view, there are only four basic filter types: *low-pass*, *high-pass*, *bandpass*, and *band rejection*. All filter configurations can be placed in one of these categories.

The characteristic curves of Fig. 11.1 show the differences among the four basic categories. The low-pass filter of Fig. 11.1a permits all frequencies below  $f_c$  to pass, but rejects all frequencies above  $f_c$ . The high-pass filter (Fig. 11.1b) rejects all frequencies below  $f_c$  but passes all others. The bandpass filter (Fig. 11.1c) allows only a limited range of frequencies to pass. The band reject filter (Fig. 11.1d) permits all frequencies, except those in a limited range, to pass.



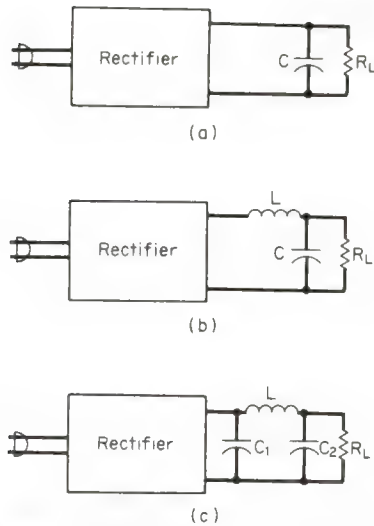
**Fig. 11.1** Characteristic curves for four basic types of filters: (a) Low-pass filter. (b) High-pass filter. (c) Bandpass filter. (d) Band-reject filter; also called bandstop filter.

Filters composed of  $R$ ,  $L$ , and  $C$  elements may be referred to as *passive filters*. Another type that utilizes a high-gain operational amplifier is the *active filter*. Besides an amplifier, an active filter also requires an external power source. The most commonly used active filter contains  $R$  and  $C$  elements.

The principal advantage of the active filter over the passive type is that inductors are not required in the former. This is especially important at the lower frequencies where inductors become excessively large and lossy.

**APPLICATIONS** The applications for filters are so numerous in electronics that only a few examples will be cited here. In terms of frequency range, filters are used at frequencies from less than a hertz in seismology to many gigahertz in microwaves.

In the rectification of alternating current to a unidirectional current it is required to filter the ripple in order to obtain an output approximating pure direct current. The simplest filter for this purpose is a capacitor across the supply output terminals, as shown in Fig. 11.2. In this application, a large-value capacitor is placed in parallel with the load resistance (Fig. 11.2a). The capacitor appears as an "open" circuit to direct current, while to alternating current it looks almost like a "short" circuit. The capacitor filter is most effective for small load currents (large values of load resistance). An example of such an application is the high-voltage power supply for a cathode-ray tube in which the anode current is typically on the order of a few hundred microamperes. As load current increases, the ripple increases and the regulation becomes poor.



**Fig. 11.2** Filters used with power supplies: (a) Pure C. (b) L section. (c) Pi.

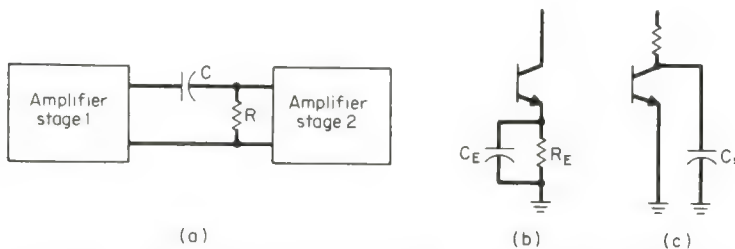
Where appreciable load current is drawn from the supply, an L-section or pi filter (Fig. 11.2*b* and *c*) is required. In the L-section filter a large value of inductance, typically 10–30 H, referred to as a *choke*, is placed in series with the load resistance.

Placing a capacitor before the L-section configuration yields the pi filter of Fig. 11.2*c*. Greater ripple attenuation and dc output voltage are realized with this circuit than with the L-section filter.

Figure 11.3 illustrates some examples of filters found in communications circuits.

The RC-coupled network of Fig. 11.3*a*, employed in cascading ac amplifier stages, behaves as a filter. Low frequencies and direct current are attenuated, while the higher frequencies pass with minimum attenuation. Another example of a filter is the capacitor bypassing the emitter resistance in Fig. 11.3*b*. This circuit also tends to attenuate the lower frequencies. In fact, the low-frequency response of an ac amplifier is largely determined by the coupling and bypass networks which behave as filters.

At higher frequencies, the shunt capacitance of the active device, such as a transistor, and shunt stray and wiring capacitances, limit the high-frequency response of the amplifier. The shunting capacitance is lumped as capacitor  $C_s$  in Fig. 11.3*c*. Capacitor  $C_s$  and its associated resistors behave as a filter that attenuates the higher frequencies. This accounts for the falling off of high frequencies in an amplifier.



**Fig. 11.3** Examples of filters found in communications circuits: (a) RC coupling. (b) Bypassing emitter resistor. (c) Shunt capacitance across amplifier output.

## 11.3 FILTER TERMINOLOGY

Before making a detailed study of filter theory, we will first review some of the more important terms that are applicable to this area.

The *configuration* of a circuit is simply the layout of its parts in relation to one another. An example is the pi filter shown in Fig. 11.2c. The parts are arranged on the schematic diagram in the form of the Greek letter  $\pi$ , and we can say that the filter has that configuration.

All filter circuits can be arranged into one of two possible configurations: the *lattice* and the *ladder*. These are illustrated in Fig. 11.4. The filter *elements* (that is, the inductors, capacitors, and resistors) are shown as separate impedances labeled  $Z$ . The most important difference between the ladder and the lattice is symmetry. The lattice,

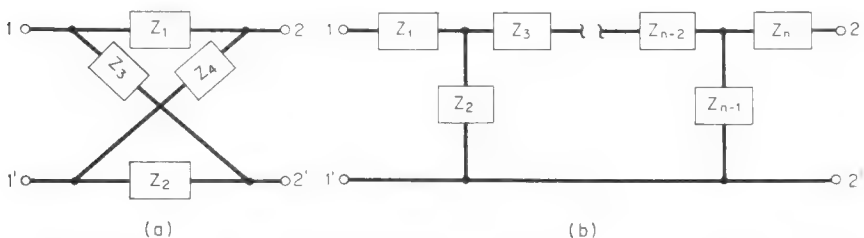


Fig. 11.4 Typical filter configurations: (a) The bridge, or lattice, circuit. (b) The ladder circuit.

or bridge, is *symmetrical*. This means that each terminal has a mirror-image counterpart on the opposite side of the circuit. An example is the equivalence of  $Z_1$  and  $Z_2$  or  $Z_3$  and  $Z_4$  in Fig. 11.4a. The ladder circuit does not have this property of symmetry. Because the lattice is symmetrical, it may drive (or be driven by) balanced loads, such as center-tapped transformers. The ladder can only drive unbalanced loads, such as a grounded resistor, tube grid, or transistor base.

The term *balance* always means with respect to ground. Figure 11.4b shows a ladder-circuit configuration. As an additional way of distinguishing between the circuits, any of the terminals of the lattice may be grounded without affecting the circuit performance while only terminals 1' and 2' of the ladder can be grounded.

Although there may be variations of these configurations, the ladder and the lattice are the basis of all filter structures. The type of element may change, i.e., capacitive, inductive, or resistive, but the configuration will not.

*Insertion loss* is the loss in signal strength that the frequencies in the passband experience in passing through the filter. (The *passband* of a filter is the range of frequencies that are supposed to be able to pass through the circuit.) If the filter were absent, and the source and load were connected directly to each other, the output signal would increase by the amount of the insertion loss.

The bandwidth of a filter is the frequency difference between two points on the filter-response curve that have a specified insertion loss. The points on the curve generally used to define bandwidth are the half-power, or  $-3$ -dB, points. These are the points where the filter response is down 3 dB from the maximum point on the curve. On the curve of Fig. 11.5, the bandwidth is between  $f_1$  and  $f_2$ , where  $f_1$  is the lower frequency at a point 3 dB down from maximum, and  $f_2$  is the upper frequency at a point 3 dB down from maximum. The formula for bandwidth is

$$\text{Bandwidth, Hz} = f_2 - f_1 \quad (11.1)$$

where  $f_1$  and  $f_2$  are at the  $-3$ -dB points.

**example 11.1** Two points on a response curve that are at the  $-3$ -dB points are at 10 and 1000 Hz for  $f_1$  and  $f_2$  respectively. What is the bandwidth?

**solution**

$$\begin{aligned} \text{Bandwidth, Hz} &= f_2 - f_1 \\ &= 1000 - 10 \text{ Hz} = 990 \text{ Hz} \end{aligned}$$

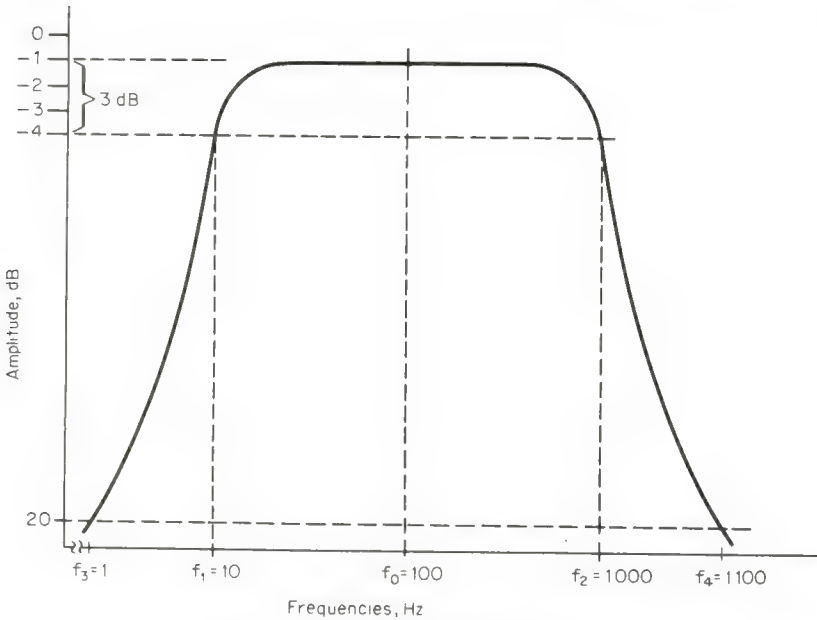


Fig. 11.5 A typical bandpass filter response.

Instead of expressing the bandwidth as the difference between two frequencies, it is sometimes expressed as a percent of frequency (*percent bandwidth*). For example, if a 100-kHz bandpass filter has a bandwidth of 10 percent, its bandwidth is 10 percent of 100 kHz, or 10 kHz. This can be expressed as  $\pm 5$  percent, or  $\pm 5$  kHz with respect to 100 kHz.

The *center frequency*  $F_0$  may be defined as either the *arithmetic* or the *geometric center*. For narrow bandwidths these are practically equal. The formulas are:

$$F_0 = \frac{f_1 + f_2}{2} = \left( \begin{array}{l} \text{Arithmetic center} \\ \text{frequency, Hz} \end{array} \right) \quad (11.2)$$

$$F_0 = \sqrt{f_1 \times f_2} = \left( \begin{array}{l} \text{Geometric center} \\ \text{frequency, Hz} \end{array} \right) \quad (11.3)$$

Using our previous example,

$$F_0 = \frac{1000 + 10}{2} = 505 \text{ Hz} \quad (\text{arithmetic center})$$

$$F_0 = \sqrt{1000 \times 10} = 100 \text{ Hz} \quad (\text{geometric center})$$

If a filter-response curve is plotted on graph paper having the abscissa (horizontal axis) marked in evenly spaced frequencies, then the arithmetic center frequency is exactly halfway between the  $-3$ -dB points. As an example, for the response curve of Fig. 11.1c the arithmetic center is 455 kHz.

If a filter-response curve is plotted on graph paper having a logarithmic abscissa, then, the geometric center frequency will be exactly halfway between the  $-3$ -dB points. For the response curve of Fig. 11.5 the abscissa is logarithmically marked. Note that the same distances are used for 10 to 100 Hz as for 100 to 1000 Hz. The geometric center on this curve is 100 Hz, which is seen to be halfway between the  $-3$ -dB points.

The arithmetic center is used only for very narrow bandwidths; then the arithmetic and geometric centers are equal. In any case, the geometric center is always correct, and when we speak of the center frequency, we shall mean the geometric center.

*Selectivity* is a measure of the effectiveness of a filter in performing its function. If



## 11-6 Filters

the filter were perfect, it would have no insertion loss in the passband and infinite attenuation at all other frequencies (the *stop band*). In practice, it takes a finite band of frequencies to change from very little attenuation to a greater amount.

One way to measure selectivity is by the *shape factor*. Shape factor is often defined as the ratio of the -3-dB bandwidth to the bandwidth where the insertion loss is 20 dB. (The -6- and -60-dB bandwidths are also used.) The closer this ratio is to a value of 1.0, the greater the selectivity of the filter. The shape factor for the response of Fig. 11.5 is

$$\text{Shape factor} = S = \frac{f_4 - f_3}{f_2 - f_1} \quad (11.4)$$

where  $f_4, f_3$  = points on the response curve that are down 20 dB from maximum  
 $f_2, f_1$  = points on the response curve that are down 3 dB from maximum

**example 11.2** Using the values of  $f_1, f_2, f_3$ , and  $f_4$  in Fig. 11.5, what is the shape factor of the selectivity curve?

**solution**

$$S = \frac{f_4 - f_3}{f_2 - f_1} = \frac{1100 - 1}{1000 - 10} = 1.11 \quad \text{Answer}$$

When lumped components are used to make a filter, there are a number of undesirable distributed elements—called *parasitic elements*—that cause attenuation of the *desired* frequencies. Parasitic elements may also change the shape of the response curve from the desired characteristic. A few examples of parasitic elements are: losses in iron cores of chokes, distributed wiring capacitance, capacitance between windings of a choke, resistance of connecting wires, and leakage inductance. Parasitic elements cannot be eliminated completely. In some cases, they may have to be considered during the design procedure.

This is not a complete list of all terms used to describe filters, but it includes the most important ones for evaluating filters.

### 11.4 FILTER COMPONENTS

Passive filters may be classed as *resistive-reactive* type or *totally reactive*. The resistive-reactive type is called *RC* or *RL*, depending whether an inductor and resistor (*RL*) or capacitor and resistor (*RC*) are used. The totally reactive filter uses capacitors and inductors, with resistance present only as an undesired distributed component. These filters are also called *LC* types because of the components utilized in their construction.

The *RL* and *RC* types are sometimes employed to generate high-pass or low-pass responses, but special modifications must be made to these in order to generate the bandpass or band-rejection response. Even then they are not as efficient as the *LC* type, owing to their inherent resistance dissipation.

### 11.5 METHODS OF DESIGNING LC FILTERS

Two different approaches are used for designing *LC* filters. One method is to treat the filter as a transmission line. This approach is called the *image-parameter design* method. For this method, the lumped components (which in *LC* filters are inductors and capacitors) are described in terms of the distributed components in a transmission line. The other approach to filter design is called the *network method of design*. (It is also called the *insertion loss method*, or *polynomial method*, or the *exact method*.) It deals directly with the filter as a circuit having parameters instead of dealing with the transmissionline.

**IMAGE-PARAMETER DESIGN** The image-parameter method of design was developed during the conduction of early experiments on telephone transmission lines. It was found that the attenuation of telephone lines increased with frequency, owing to the large amount of distributed capacitance, especially in the early types of lines. Inserting inductors reduced the attenuation and made it fairly constant over a range of

frequencies. It was also noted that the attenuation increased very quickly at a particular frequency called the *cutoff frequency*.

As the requirements for bandwidth in the transmission line increased, various schemes were used to simultaneously increase bandwidth and keep the attenuation low. As a result of experiments, image-parameter filters were developed. When the equivalent circuits of loaded transmission lines were constructed with lumped components, it was found that they had different characteristics from those of the loaded line. The bandwidth and attenuation were not the same as in the transmission line because the filter is only an approximation of a transmission-line circuit. This is one of the difficulties in designing such filters.

The image-parameter filter has evolved into two basic types: *the constant-k type*, and *the m-derived type*. The constant-k type is the result of the first attempt to build an artificial transmission line. When component values of the constant-k type are modified in a certain way so as to vary the impedance and the shape factor, the result is an m-derived filter.

**NETWORK METHOD OF DESIGN** Modern mathematics and the advent of the computer have done much to change filter design. Tables and nomographs obtained through the use of computers make it possible to obtain a desired filter characteristic with relative ease. Using only simple mathematics, it is possible to design a filter for almost any application and shape factor.

The image-parameter method of designing a filter is based on transmission lines. It attempts to simulate a transmission line with inductors and capacitors, that approximates the distributed constants in a coaxial cable, and obtain a certain impedance and frequency response simultaneously. For a long time it has been known that filters could be designed from response-curve formulas. The difficulty was that the equations were so long they were nearly impossible to solve. A computer was required for their solution, but at that time computers were not available for such applications.

When more and more computers became available, the required equations were solved, and tables of elements were generated that could be used in many different filter designs. It was desirable to make the tables useful for *any* frequency or impedance level; so they were developed in such a way that this information could be added later. Such tables are said to be *normalized*. The values are referenced to 1  $\Omega$  and  $2\pi$  Hz (1 rad/s), meaning that the impedance and the cutoff frequency are 1  $\Omega$  and  $2\pi$  Hz, respectively, for the filter elements listed. This is equivalent to the mechanical operation of *scaling* an object. To change the size of a scaled object, you simply divide or multiply by the number of times bigger or smaller that the object is to be. For example, if a scale model of a boat is  $\frac{1}{25}$  of the actual size, then each and every dimension on the model must be multiplied by 25 to obtain its actual size. In a similar manner, the tables used for the network method of design can be considered to be scaled down. In order to get the desired values, it is necessary to multiply by the desired impedance in ohms and the desired frequency in hertz. By making the tables this way, they apply to *any* impedance and frequency.

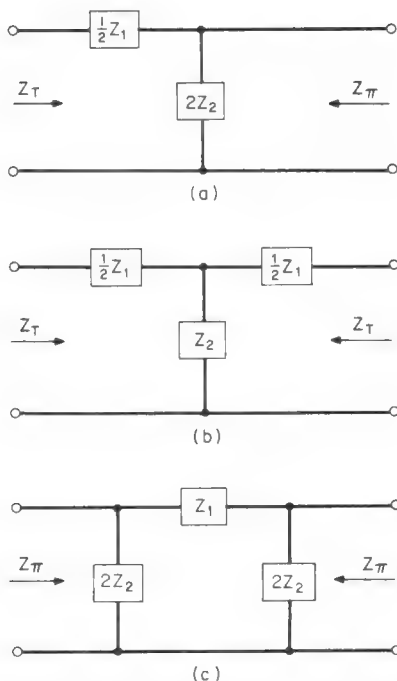
We will give examples of each method of design, starting with the image-parameter approach. The constant-k and m-derived filters are designed by this method.

## 11.6 THE SIGNIFICANCE OF FILTER IMPEDANCE

Constant-k filters can be made in any of the three configurations shown in Fig. 11.6. Each of these is a form of ladder network, and is identified according to the alphabet letter that it forms. Thus we have an L, a T, and a  $\Pi$  section. A number of these sections can be added to obtain a sharper cutoff (or other desired characteristic).

An important property of constant-k filters—and for that matter, *all* types of filters—is their impedance. This is because of the *maximum power-transfer theorem*, which states that the maximum possible amount of power that can be delivered to a load occurs when the load impedance “matches” the internal impedance of the generator. For ac generators, the match must be achieved by using a *conjugate impedance*. Thus, if the internal impedance of the generator is  $7 + j11 \Omega$ , then maximum power will be delivered to the load impedance when its value is  $7 - j11 \Omega$ .

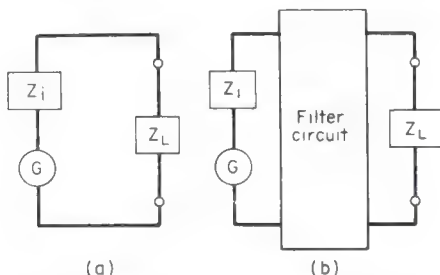
Figure 11.7 shows the effect of the filter on the impedance match. Suppose we have



**Fig. 11.6** Configurations of the basic filter sections: (a) An L section. (b) A T section. (c) A pi section.

a circuit like that of Fig. 11.7a, in which the load impedance  $Z_L$  is matched to the internal impedance  $Z_i$  of the generator  $G$ . In accordance with the maximum power-transfer theorem, the greatest amount of power (for this generator) is being delivered to the load. Now if we insert a filter circuit between the generator and load, as shown in Fig. 11.7b, then the presence of the filter can result in an improper match of impedances. (It is important to note that inserting a transmission line between a generator and a load can also produce a mismatch, and therefore transmission-line impedance and filter impedance are both important, and have a similar effect on the circuit into which they are inserted.)

Suppose that the internal impedance and the load impedance of the circuit in Fig.



**Fig. 11.7** Illustrating the importance of filter circuit impedance: (a) A circuit designed for maximum power transfer. (b) Unless properly designed, the filter circuit could unbalance the impedance match.

11.7a are conjugates, and they both have the same impedance—say  $500\ \Omega$ . If the input impedance to the filter circuit is also  $500\ \Omega$ —looking from either end—then, no impedance mismatch will occur between the generator and the load.

Now suppose the internal impedance of the generator is different from the load impedance. If the filter is designed in such a way that the generator operates into the correct impedance, and the load operates into the correct impedance, then the filter is actually serving two functions: elimination of unwanted frequencies and matching of impedances. Whenever a filter is designed in such a way that the generator “sees” the conjugate of its own internal impedance when looking into the filter output terminals, then the filter is said to be operating on an *image impedance basis*.

The *characteristic impedance* of a filter, and also of a transmission line, is the value of load impedance that must be connected across the output terminals in order to match the load to the filter (or line). This is also known as the *surge impedance*. If a filter, or transmission line, is terminated in a resistance that is equal to its impedance, then all the energy moving through the filter (or through the line) will be dissipated by that resistance.

The characteristic impedance  $Z_0$  of a filter or transmission line is given by the equation

$$Z_0 = \sqrt{Z_{sc}Z_{oc}} \quad (11.5)$$

where  $Z_{sc}$  = input impedance with the output terminals short-circuited together  
 $Z_{oc}$  = input impedance with the output terminals open

## 11.7 DESIGN OF CONSTANT-k FILTERS

The application of the constant- $k$ , as well as other  $LC$  filters, is generally practical over a range of frequency up to 100 MHz. Their applications in communication circuits include the limiting of the signal spectrum to a receiver (preselector), separating or combining the spectrum so that parts of it may be individually processed (duplexers and diversity receiver combiners), separating r-f signals, and reducing interference.

Although constant- $k$  filters are easy to design and make, they have a number of shortcomings. Their attenuation is low near its cutoff frequency, and their image impedance is not constant with frequency. Also, the phase shift in the passband is not proportional to frequency.

It is apparent when we examine the circuits of Fig. 11.6 that the impedance of a filter circuit will normally change with frequency. As an example, assume that  $Z_1$  is an inductor and that both impedances marked  $2Z_2$  are capacitors in the circuit of Fig. 11.6c. If the input frequency increases, then the inductive reactance increases, and at the same time the capacitive reactance decreases. The result is a change in filter impedance. reactance decreases. The result is a change in filter impedance.

If the inductance and capacitance are chosen so that the increase in inductive reactance is exactly offset by the decrease in capacitive reactance, then their product will be a constant, which is called the constant  $k$ . Expressed mathematically,

$$k = \sqrt{Z_1 Z_2} \quad (11.6)$$

where

$k$  = an impedance independent of frequency  
 $Z_1, Z_2$  = series and shunt impedances of Fig. 11.6

The attenuation of all frequencies will be a constant up to the cutoff frequency. If the  $Q$  of the reactive components in the filter is high, then the cutoff will be quite sharp. As the  $Q$  is made lower and lower, the reduction in attenuation beyond cutoff becomes more and more gradual.

**LOW-PASS FILTER DESIGN** Figure 11.8 shows three possible configurations for a constant- $k$  low-pass filter. To design such a filter, it is necessary to know the cutoff frequency  $f_c$ —that is, the frequency the filter is to pass. This is designated  $f_c$  in Fig. 11.9. It is also necessary to know the impedance of the load to be connected across the output terminals of the filter. This load, which is designated as  $Z_o$ , must be the conjugate of the characteristic impedance of the filter in order for maximum power transfer to occur. Looking in from the output terminals, this is also the image impedance.

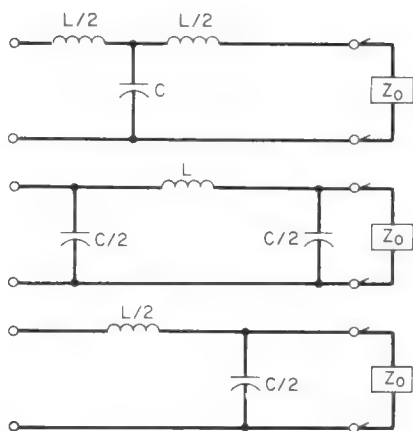


Fig. 11.18 Three forms of constant- $k$  low-pass filters.

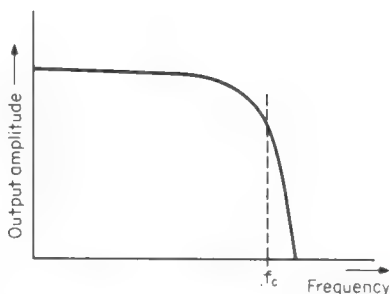


Fig. 11.19 Characteristic curve of a constant- $k$  low-pass filter.

By knowing the value of  $f_c$  and  $Z_0$ , it is possible to design the filter from the following equations which are given here without derivation:

$$L = \frac{Z_0}{3.14f_c} \quad (11.7)$$

$$C = \frac{1}{3.14f_c Z_0} \quad (11.8)$$

where  $L$  = inductance, H

$C$  = capacitance, F

$Z_0$  = load impedance,  $\Omega$

Once we have calculated the required values of  $L$  and  $C$ , we can verify that this is a constant- $k$  filter by substituting the values into the equation

$$Z_0 = k = \sqrt{\frac{L}{C}} \quad (11.9)$$

where  $Z_0$ ,  $L$ ,  $C$  = values in Eqs. (11.8) and (11.7)

$k$  = constant  $k$ , and is equal to the image impedance for all frequencies that the filter is designed to pass

The following example illustrates the design of a low-pass constant- $k$  filter. The cutoff frequency is given as 1000 Hz. All frequencies above 1000 Hz are attenuated. A filter of this characteristic might be used in narrow-band transmission of information.

**example 11.3** Design a constant- $k$  low-pass filter that will attenuate frequencies above 1000 Hz ( $f_c = 1000$  Hz). The load across the filter output terminals,  $Z_0$ , is a pure resistance of 600  $\Omega$ . In order to obtain sharp cutoff characteristics, use a three-section ladder configuration.

**solution** The greater the number of filter sections used, the sharper the cutoff—hence the three-section ladder. Not all applications could justify the cost of components, and in many cases it is not necessary that the cutoff be sharp.

Figure 11.10 shows the circuit. The three-section filter can be made by combining two  $L$  sections with a  $T$  section, as shown in Fig. 11.10a. The first step is to calculate the required value of  $L$  from Eq. (11.7).

$$\begin{aligned} L &= \frac{Z_0}{3.14f_c} \\ &= \frac{600}{3.14 \times 1000} = 0.191 \text{ H} \end{aligned} \quad (11.7)$$



As shown in Fig. 11.10a, each inductor in this illustration actually has a value of  $L/2$ , or 0.095 H.

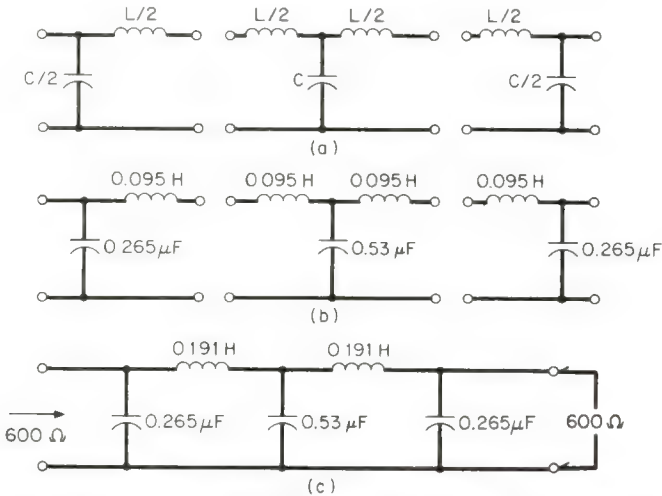
The next step is to calculate the required value of  $C$  from Eq. (11.8).

$$C = \frac{1}{3.14 f_c Z_o}$$

$$= \frac{1}{3.14 \times 1000 \text{ Hz} \times 600 \Omega} = 0.53 \mu\text{F} \quad (11.8)$$

Each capacitor in Fig. 11.10a actually has a value of  $C/2$ , or  $0.265 \mu\text{F}$ .

Figure 11.10b shows the circuit with the calculated component values. When the sections are combined to produce the required ladder filter, the two halves of  $L$  are combined to produce  $L(L/2 + L/2 = L)$  as shown in Fig. 11.10c.



**Fig. 11.10** An example of a constant- $k$  filter design: (a) Prototype. (b) Element values. (c) Composite filter design.

If this is truly a constant- $k$  filter, then the characteristic impedance will be  $600 \Omega$ , as calculated from Eq. (11.9).

$$Z_o = k = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.191}{0.53 \times 10^{-6}}}$$

$$\approx \sqrt{360\,000} \approx 600 \Omega$$

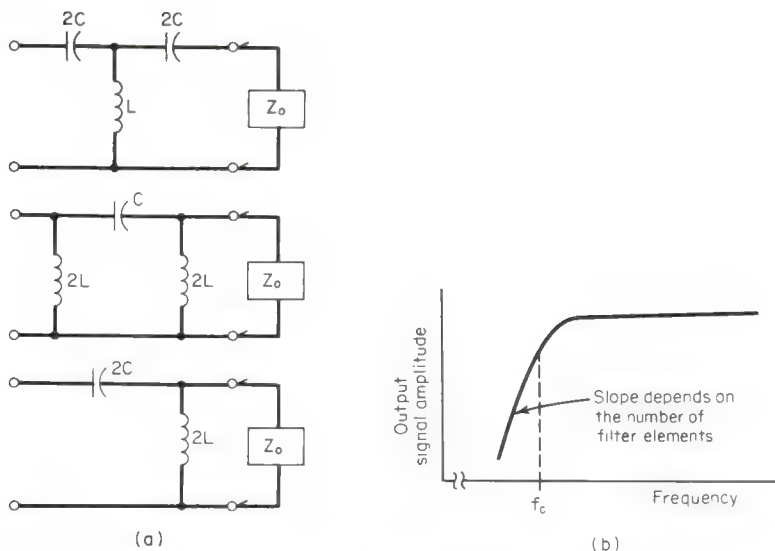
It will not be possible to obtain capacitors and inductors with the precise computed values for the filter. Any compromise in component values will also result in a compromise in shape factor and other characteristics.

The impedance of the filter must be identical with the load impedance because the circuit is the same when viewed from either end. In practice, an improved impedance match to the source and load can be obtained by adding an  $m$ -derived half-section at each end of the constant- $k$  filter of Fig. 11.10c. These added half-sections would be designed to match the image impedance of the filter to the resistive terminations of the source and the load.

**HIGH-PASS CONSTANT- $k$  FILTER DESIGN** The arrangement of components for a high-pass constant- $k$  filter design and the resulting characteristic curve are shown in Fig. 11.11. Figure 11.11a shows the three possible configurations for a constant- $k$  high-pass filter. As with the low-pass filters, these configurations are named according to the letter of the alphabet that they are similar to: T,  $\pi$ , and L. The equations for calculating the value of  $L$  and  $C$  are given here without derivation.



## 11-12 Filters



**Fig. 11.11** Circuits and characteristic curves of constant- $k$  high-pass filter: (a) Three forms of constant- $k$  high-pass filter circuits. (b) Characteristic curve.

$$L = \frac{Z_o}{12.6f_c} \quad (11.10)$$

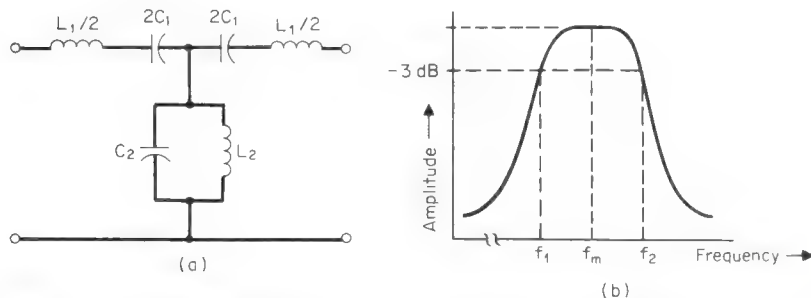
$$C = \frac{1}{12.6f_c Z_o} \quad (11.11)$$

where  $L$  = inductance, H  
 $C$  = capacitance, F  
 $Z_o$  = load impedance,  $\Omega$

Components  $L$  and  $C$  are identified in the circuit of Fig. 11.11a. The characteristic impedance is  $Z_o$ , which is also the value of  $k$ , and also the value of the load impedance. The cutoff frequency is  $f_c$  and its meaning is indicated in Fig. 11.11b.

**BANDPASS FILTER DESIGN** The constant- $k$  bandpass filter circuit and its characteristic curve are shown in Fig. 11.12.

The series circuits in Fig. 11.12a (comprised of  $L_1$  and  $C_1$ ) are resonant to the frequency being passed. Actually, when  $L_1$  and  $C_1$  are tuned to the center frequency of the passband, there is very little opposition at that frequency. The parallel circuit



**Fig. 11.12** Bandpass filter and its characteristic curve: (a) Constant- $k$  bandpass filter. (b) Bandpass filter characteristic curve.

(comprised of  $L_2$  and  $C_2$ ) is virtually an open circuit to the center frequency—which is also the parallel resonant frequency.

The characteristic curve of this filter circuit, shown in Fig. 11.12*b*, also defines the frequencies used in the calculations of inductance and capacitance. The equations for  $L_1$ ,  $L_2$ ,  $C_1$ , and  $C_2$  are given here without derivation.

$$L_1 = \frac{Z_o}{3.14(f_2 - f_1)} \quad (11.12)$$

$$L_2 = \frac{(f_2 - f_1)Z_o}{12.6f_1f_2} \quad (11.13)$$

$$C_1 = \frac{f_2 - f_1}{12.6f_1f_2Z_o} \quad (11.14)$$

$$C_2 = \frac{1}{3.14(f_2 - f_1)Z_o} \quad (11.15)$$

where  $L_1, L_2$  = are identified in Fig. 11.12*a* and are measured in henrys

$C_1, C_2$  = are identified in Fig. 11.12*a* and are measured in farads

$f_1, f_2$  = limits of bandwidth at the -3-dB points

$Z_o$  = characteristic impedance,  $\Omega$

The constant  $k$  is equal to the image impedance, and is defined by the equation

$$k = \sqrt{\frac{L_1}{C_2}} \quad (11.16a)$$

$$= \sqrt{\frac{L_2}{C_1}} = Z_o \quad (11.16b)$$

The geometric center frequency  $f_m$  is given by the equation

$$\begin{aligned} f_m &= \sqrt{f_1f_2} \\ f_m &= \frac{1}{6.28\sqrt{L_1C_1}} \\ &= \frac{1}{6.28\sqrt{L_2C_2}} \end{aligned} \quad (11.17)$$

In the next example, the design of a bandpass filter is illustrated. Such a filter may find application in a spectrum analyzer, where a band of frequencies is to be viewed at the exclusion of all other frequencies.

**example 11.4** (a) Design a bandpass filter for passing frequencies between 90 and 110 kHz. The load is 1000  $\Omega$ . (b) Calculate its center frequency.

**solution** (a) The bandpass filter circuit is shown in Fig. 11.12*a*. The load impedance is 1000  $\Omega$ , and this must equal the characteristic impedance  $Z_o$ . The lower frequency limit  $f_1$  is 90 kHz, and the upper frequency limit is 110 kHz. With these values, the required inductances,  $L_1$  and  $L_2$ , and the capacitance values,  $C_1$  and  $C_2$ , can be calculated.

$$\begin{aligned} L_1 &= \frac{Z_o}{3.14(f_2 - f_1)} \\ &= \frac{1000 \Omega}{3.14(110 \text{ kHz} - 90 \text{ kHz})} = 15.9 \text{ mH} \end{aligned} \quad (11.12)$$

$$\begin{aligned} L_2 &= \frac{(f_2 - f_1)Z_o}{12.6f_1f_2} \\ &= \frac{(110 \text{ kHz} - 90 \text{ kHz})(1000 \Omega)}{12.6(90 \text{ kHz})(110 \text{ kHz})} = 160 \mu\text{H} \end{aligned} \quad (11.13)$$

$$\begin{aligned} C_1 &= \frac{f_2 - f_1}{12.6f_1f_2Z_o} \\ &= \frac{110 \text{ kHz} - 90 \text{ kHz}}{12.6(90 \text{ kHz})(110 \text{ kHz})(1 \text{ k}\Omega)} = 160 \text{ pF} \end{aligned} \quad (11.14)$$

$$\begin{aligned}
 C_2 &= \frac{1}{3.14(f_2 - f_1)Z_0} \\
 &= \frac{1}{3.14(20 \text{ kHz})(1 \text{ k}\Omega)} = 0.0159 \mu\text{F}
 \end{aligned} \quad (11.15)$$

If these values are correct, then the value of the constant  $k$  will be equal to the characteristic impedance of the filter. This can be checked with Eq. (11.16a):

$$\begin{aligned}
 k &= \sqrt{\frac{L_1}{C_2}} = \sqrt{\frac{15.9 \times 10^{-3}}{0.0159 \times 10^{-6}}} \\
 k &= Z_0 = 1000 \Omega
 \end{aligned}$$

The value of  $k$  can also be computed from Eq. (11.16b):

$$\begin{aligned}
 k &= \sqrt{\frac{L_2}{C_1}} = \sqrt{\frac{160 \times 10^{-6}}{160 \times 10^{-12}}} \\
 k &= Z_0 = 1000 \Omega
 \end{aligned}$$

The value of  $k$  has been calculated in two different ways, using  $L_1$  and  $C_2$ , and also  $L_2$  and  $C_1$ . In both cases it is equal to the characteristic impedance required to match the filter to the load. This, in turn, is the image impedance as seen by the load.

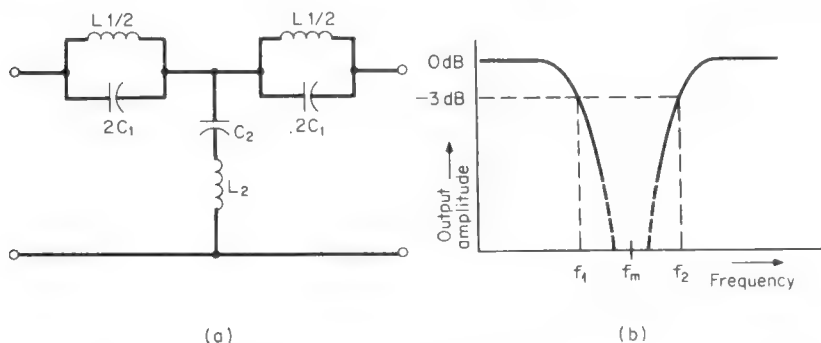
(b) The center frequency of the band of frequencies passed by the filter is obtained from Eq. (11.3).

$$F_0 = \sqrt{f_1 \times f_2} = \sqrt{90 \times 110 \text{ kHz}} = 99.5 \text{ kHz}$$

This is the geometric center frequency which is sometimes designated  $f_m$ , the subscript  $m$  standing for midfrequency. The arithmetic frequency, when calculated from Eq. (11.2) is 100 kHz, which is close to the geometric mean.

**BAND-REJECTION FILTER DESIGN** Band-rejection filters are sometimes called *band-stop filters*, or *notch filters*.

Figure 11.13 shows a constant- $k$  band-rejection filter circuit and its characteristic curve. The parallel-tuned circuits,  $L_1$  and  $C_1$ , allow all frequencies to pass through the



**Fig. 11.13** Band-stop filter and its characteristic curve: (a) Band-rejection filter. (b) Filter response.

filter except the frequencies near resonance. The resonant frequency of these tuned circuits is the center frequency of the band being rejected. The series-resonant circuit,  $L_2$  and  $C_2$ , short-circuits the unwanted frequency to ground. The width of the band being rejected is a direct function of the circuit  $Q$ . The slope of the curve—that is, the sharpness of tuning—is increased by adding filter sections.

The equations for  $L_1$ ,  $L_2$ ,  $C_1$ , and  $C_2$  are given here without derivation.

$$L_1 = \frac{(f_2 - f_1)Z_0}{3.14f_1f_2} \quad (11.18)$$

$$L_2 = \frac{Z_0}{12.6(f_2 - f_1)} \quad (11.19)$$

$$C_1 = \frac{1}{12.6(f_2 - f_1)Z_0} \quad (11.20)$$

$$C_2 = \frac{f_2 - f_1}{3.14f_1f_2Z_0} \quad (11.21)$$

where  $L_1, L_2$  = are identified in Fig. 11.13a and are measured in henrys

$C_1, C_2$  = are identified in Fig. 11.13a and are measured in farads

$f_1, f_2$  = limits of the bandwidth being rejected as measured at the 3-dB points

$Z_0$  = characteristic impedance,  $\Omega$

The constant  $k$  is equal to the image impedance and is defined by the equation

$$k = \sqrt{\frac{L_1}{C_2}} \quad (11.22a)$$

$$k = \sqrt{\frac{L_2}{C_1}} = Z_0 \quad (11.22b)$$

The geometric center frequency  $f_m$  is given by the equation

$$\begin{aligned} f_m &= \sqrt{f_1 f_2} \\ &= \frac{1}{6.28 \sqrt{L_1 C_1}} = \frac{1}{6.28 \sqrt{L_2 C_2}} \end{aligned} \quad (11.23)$$

## 11.8 SUMMARY OF CONSTANT- $k$ FILTER CHARACTERISTICS

In each of the relationships given for the design of constant- $k$  filters, it was necessary to know the value of the load impedance to be placed across the filter output terminals. The values of inductances and capacitances are always dependent upon (among other things) this impedance value. Use of the constant  $k$  was not necessary, and in our examples the only use of constant  $k$  was to show that it is equal to the image impedance. However, the derivations of the equations for  $L$  and  $C$ , which are rather tedious and were not given, do depend on use of the constant  $k$ .

The shapes of the characteristic curves for the constant- $k$  filters cannot be controlled by the design engineer. He can, of course, add more sections to increase the selectivity, but the additional cost may not be justified. The sides of the selectivity curves of bandpass and band-stop filters are referred to as the *skirts*, and the shapes of these curves—as measured by the shape factor—are often referred to as the *skirt selectivity*. This is simply an indication of how sharply the circuit tunes, as is the shape factor. Constant- $k$  filters are characterized by poor skirt selectivity when compared to other types of filters.

As we shall see in the next section, the first step in designing an  $m$ -derived filter is to design a constant- $k$  filter. Then, the constant- $k$  filter is modified by adding a reactance to either the series or shunt leg. Since the constant- $k$  filter is designed first, it is often referred to as the *prototype*.

## 11.9 m-DERIVED FILTERS

Besides having improved skirt selectivity, the frequency of infinite attenuation can be readily selected in the  $m$ -derived filter. It still shares with the constant- $k$  filter the disadvantage of not having a constant image impedance throughout its passband. The combination of constant- $k$  and  $m$  sections, however, provides a flexibility in filter design not realized with the constant- $k$  filter alone.

The constant- $k$  filter has a rather gradual *roll-off*: that is, the amplitude of the signal does not change rapidly at the skirts. There are some applications where a sharper response is needed. This can be accomplished by adding reactance to either the series or parallel (shunt) branch. Figure 11.14 shows three examples with a reactance added to the shunt path. Note that the same configurations ( $T$ ,  $\pi$ , and  $L$ ) are employed as were used for constant- $k$  filters. In fact, these are identical with the constant- $k$  low-pass filters except that an inductor has been added to resonate with the shunt capacitance. The series elements  $L_1$  are unchanged, and, therefore, these configurations are referred to as series-type  $m$ -derived filters. The  $m$ -derived filter, then, is

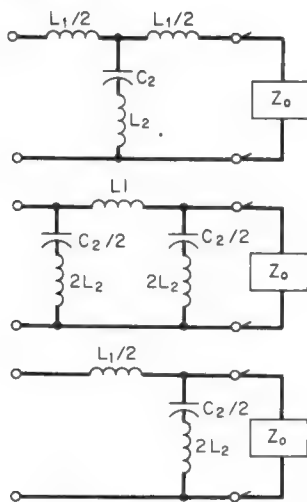


Fig. 11.14 Low-pass series-type  $m$ -derived filter configurations.

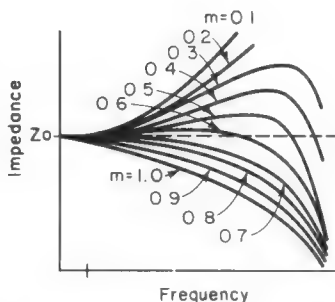


Fig. 11.15 The effect of  $m$  on filter impedance.

named for the branch that is *unchanged* from the constant- $k$  type. If the shunt paths  $C_2$  are unchanged, and reactive components are added to the series elements, the configuration is then called a shunt-type  $m$ -derived filter.

Returning again to the circuits of Fig. 11.14, we see that the shunt path is now a resonant circuit. There will be a resonant frequency at which the circuit presents a short across the line. That frequency, called the *infinite attenuation frequency*  $f_x$ , and the cutoff frequency  $f_c$  do not have the same value. Cutoff occurs at the point where the output signal amplitude is down 3 dB from the maximum value, just as in the constant- $k$  filter.

The value of  $m$  in the  $m$ -derived filter is always less than 1, and may be obtained mathematically from one of the two equations given below.

For low-pass,  $m$ -derived filters:

$$m = \sqrt{1 - \frac{f_x^2}{(f_c)^2}} \quad (11.24a)$$

For high-pass,  $m$ -derived filters:

$$m = \sqrt{1 - \frac{f_c^2}{f_x^2}} \quad (11.24b)$$

where  $f_c$  = cutoff frequency Hz

$f_x$  = infinite attenuation frequency, Hz

$m$  = a constant used in determining the values of elements

A special case occurs for  $m = 1.0$ . If the value of 1.0 is substituted into the equations for  $m$ -derived filters, the filter is converted into a constant- $k$  filter. This means that the constant- $k$  filter may be considered to be a special type of  $m$ -derived filter for which  $m = 1.0$ .

Choice of the value of  $m$  is not entirely dependent on the relationship between  $f_c$  and  $f_x$ , as might be supposed from Eqs. (11.24a) and (11.24b). Different values of  $m$  cause the filter to have different impedance characteristics. The relationship between the value of  $m$  and the value of filter impedance is illustrated in the graph of Fig. 11.15. In fact, this is one of the important advantages of  $m$ -derived filters: the designer can choose the particular impedance relationship that he wants. In the constant- $k$  filter the designer must accept the particular impedance relationship for that type of filter.

In Fig. 11.15, the constant- $k$  filter impedance is represented by the curve marked  $m = 1.0$ .

In the graph of Fig. 11.15 the curve marked  $m = 0.6$  provides the impedance versus frequency relationship which is nearest to being ideal. This curve is considered to be the best choice for many applications, and, therefore,  $m = 0.6$  is a value that is frequently chosen in practical filter design. Of course, the greater the range of frequencies over which the impedance is constant, the greater the range of frequencies over which maximum power transfer can occur.

**LOW-PASS m-DERIVED FILTER DESIGN (SERIES TYPE)** The equations for  $L_1$ ,  $L_2$ , and  $C_2$  are given here without derivation. In each case, it may be necessary to alter the calculated value in accordance with the specifications marked on Fig. 11.14. For example, suppose you are solving for  $L_1$  in the T configuration, and you determine that  $L_1 = 10$  mH from the equations below. According to the illustration, this value is divided by 2, so that the actual value of each of the two inductors will be  $L_1/2 = 10/2 = 5$  mH.

$$L_1 = m \frac{Z_0}{3.14f_c} \quad (11.25)$$

$$L_2 = \frac{1 - m^2}{4m} \times \frac{Z_0}{3.14f_c} \quad (11.26)$$

$$C_2 = m \frac{1}{3.14Z_0f_c} \quad (11.27)$$

where  $L_1$ ,  $L_2$  = series and shunt inductances, respectively, H

$C_2$  = shunt capacitance, F

$m$  = a constant obtained from the use of Eq. (11.24a)

$f_c$  = cutoff frequency, Hz

$Z_0$  = characteristic impedance of the filter circuit,  $\Omega$

If the value of  $m$  is chosen as unity ( $m = 1.0$ ), then the three equations become

$$L_1 = \frac{Z_0}{3.14f_c}$$

$$L_2 = 0$$

$$C_2 = \frac{1}{3.14Z_0f_c}$$

The results mean that no inductance is used for the shunt leg ( $L_2 = 0$ ), and the values of  $L_1$  and  $C_2$  are identical with those of the constant- $k$  filter as given by Eqs. (11.7) and (11.8). This proves that a constant- $k$  filter is a special case of an  $m$ -derived filter for which  $m = 1.0$ .

**LOW-PASS m-DERIVED FILTER DESIGN (SHUNT TYPE)** Instead of the reactive component being placed in the shunt leg (as was done for the series-type  $m$ -derived filter), it may be placed in the series leg, as shown in Fig. 11.16. In this case a parallel circuit is formed by  $L_1$  and  $C_1$  which offers high impedance at the frequency to which they are tuned. If the inductor and capacitor were perfect elements—that is, if they contained no resistance—then the opposition to signal flow at the resonant frequency of  $L_1$  and  $C_1$  would be infinite. For this reason, the resonant frequency is called the infinite attenuation frequency, as it was in the series type of Fig. 11.14.

The equations for  $L_1$ ,  $C_1$ , and  $C_2$  are given here without derivation. As before, it may be necessary to modify the values of the elements as for the series type. Thus,  $L_1$  becomes  $L_1/2$  in the T type,  $C_2$  becomes  $C_2/2$  in the  $\pi$  type, etc.

$$L_1 = m \frac{Z_0}{3.14f_c} \quad (11.28)$$

$$C_1 = \left( \frac{1 - m^2}{4m} \right) \left( \frac{1}{3.14Z_0f_c} \right) \quad (11.29)$$



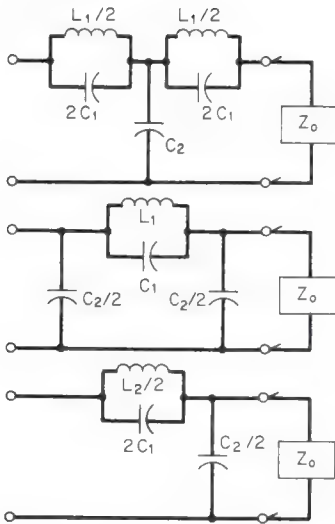


Fig. 11.16 Low-pass shunt-type  $m$ -derived filter configurations.

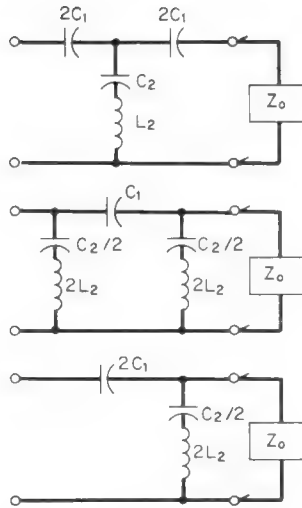


Fig. 11.17 High-pass series-type  $m$ -derived filter configurations.

$$C_2 = m \frac{1}{3.14 Z_o f_c} \quad (11.30)$$

where  $L_1$  = inductance in the series leg, H  
 $C_1$  = capacitor that resonates with  $L_1$ , F  
 $C_2$  = capacitance in the shunt leg  
 $f_c$  = cutoff frequency, Hz  
 $Z_o$  = characteristic impedance of the filter,  $\Omega$

**HIGH-PASS  $m$ -DERIVED FILTER DESIGN (SERIES TYPE)** High-pass  $m$ -derived filters can assume the same three basic configurations as are used for low-pass filters ( $T$ ,  $\pi$ , and  $L$ ). Also, each of these configurations may be designed as a series type or shunt type. Like the low-pass filter, a high-pass  $m$ -derived filter is a series type when the shunt leg is modified, and it is a shunt type when the series leg is modified. Figure 11.17 shows the three configurations for the series-type high-pass  $m$ -derived filters. The equations for  $C_1$ ,  $C_2$ , and  $L_2$  are given without derivation.

$$C_1 = \frac{1}{12.6 f_c Z_o m} \quad (11.31)$$

$$C_2 = \left( \frac{4m}{1-m^2} \right) \left( \frac{1}{12.6 f_c Z_o} \right) \quad (11.32)$$

$$L_2 = \frac{Z_o}{12.6 f_c m} \quad (11.33)$$

where  $C_1$ ,  $C_2$  = series and shunt capacitance values, F  
 $L_2$  = inductance in the shunt leg that resonates with  $C_2$ , H  
 $Z_o$  = characteristic impedance of the filter circuit,  $\Omega$   
 $f_c$  = center frequency, Hz  
 $m$  = a constant that determines the rate at which the filter characteristic rolls off.

In the circuits of Fig. 11.17 the series-resonant circuit comprised of  $C_2$  and  $L_2$  provides a short circuit to the frequency of infinite attenuation, thus preventing any signal at this frequency from passing through the filter.

**HIGH-PASS  $m$ -DERIVED FILTER DESIGN (SHUNT TYPE)** Figure 11.18 shows the three configurations for an  $m$ -derived shunt-type high-pass filter. In this case a parallel-tuned circuit comprised of  $L_1$  and  $C_1$  is tuned to the frequency of infinite impedance. The parallel-tuned circuit prevents any signal at that frequency from passing. The equations for the values of elements  $C_1$ ,  $L_1$ , and  $L_2$  are given here without derivation.

$$C_1 = \frac{1}{12.6 f_c Z_0} \quad (11.34)$$

$$L_1 = \left( \frac{4m}{1-m^2} \right) \left( \frac{Z_0}{12.6 f_c} \right) \quad (11.35)$$

$$L_2 = \frac{Z_0}{12.6 f_c m} \quad (11.36)$$

where  $L_1, L_2$  = series and shunt inductances, respectively, H

$C_1$  = capacitance that resonates with  $L_1$ , F

$f_c$  = cutoff frequency, Hz

$Z_0$  = characteristic impedance of the filter circuit,  $\Omega$

$m$  = constant arrived at in the same manner as in other  $m$ -derived filters

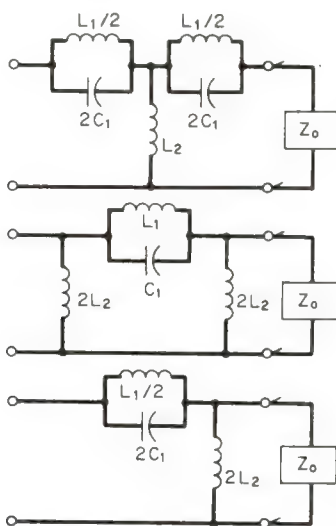
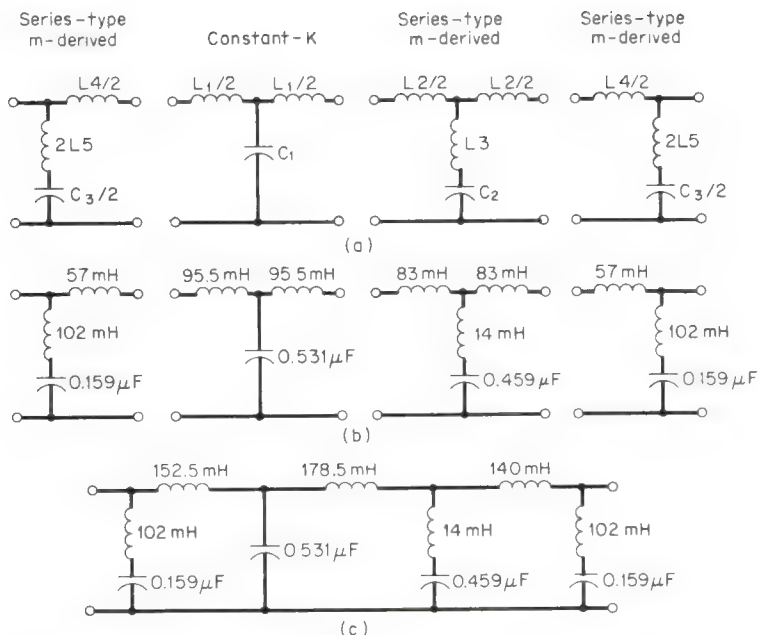


Fig. 11.18 High-pass shunt-type  $m$ -derived filter.

## 11.10 SUMMARY AND DESIGN EXAMPLE OF $m$ -DERIVED FILTERS

A number of  $m$ -derived filter sections can be combined to form a ladder-type filter. In some cases an  $m$ -derived section may be combined with a constant- $k$  section to obtain a particular overall characteristic. At the input and output ends of the filter,  $L$  sections may be used to obtain an impedance match.

Before an example of  $m$ -derived filter design is given, a few of the more important facts about  $m$ -derived filters will be summarized. The first step in designing an  $m$ -derived filter is to design the constant- $k$  prototype. A value of  $m$  is decided upon next. If the circuit is to match a load impedance over a range of frequencies, then a value of  $m = 0.6$  is generally used. The smaller the value of  $m$ , the sharper the cutoff characteristic of the filter and the greater the change in impedance with frequency. A constant- $k$  filter section added to the filter will ensure that there is attenuation beyond the cutoff point of the filter.



**Fig. 11.19** An example of  $m$ -derived filter design: (a) Filter prototype. (b) Filter sections with calculated values. (c) Composite filter.

**example 11.5** Design an  $m$ -derived ladder-type low-pass filter. The cutoff frequency  $f_c$  is to be 1000 Hz and the load impedance is to be 600  $\Omega$ .

**solution** To get a sharp cutoff, we shall use two full T sections and two L sections. See Fig. 11.19. Figure 11.19a shows the circuit. The input and output L sections are series-type  $m$ -derived filters. The first T section (reading left-to-right) is a simple constant- $k$  filter, and the other T section is an  $m$ -derived series section.

The first step is to find the value of  $L_1$  and  $C_1$  for the constant- $k$  filter.

$$L_1 = \frac{Z_0}{3.14f_c} = \frac{600}{3.14 \times 1000} \quad (11.7)$$

$$= 0.191 \text{ H (or } 191 \text{ mH)}$$

$$C_1 = \frac{1}{3.14f_c Z_0}$$

$$= \frac{1}{3.14 \times 1000 \times 600} = 0.531 \mu\text{F} \quad (11.8)$$

The values of  $L_1$  and  $C_1$  for the constant- $k$  filter are shown in Fig. 11.19b. As required by the specification of Fig. 11.19a, the value of  $L_1$  has been divided by two. The next step is to determine the values of  $L_2$ ,  $L_3$ , and  $C_2$  of the  $m$ -derived T section. The subscripts are changed from the original equations so that the various sections will not be confused.

$$L_2 = m \frac{Z_0}{3.14f_c} \quad (11-25)$$

in which the term  $(Z_0/3.14f_c)$  is equal to the constant- $k$  prototype value of  $L_1$ .

A value of  $m = 0.866$  will be used for the  $m$ -derived T section in order to get adequate attenuation. A value of  $m = 0.6$  will be used for the end sections in order to get an impedance match over a range of frequencies.

$$L_2 = \frac{0.866 \times 600}{3.14 \times 1000} = 0.166 \text{ H (or } 166 \text{ mH)}$$

$$\begin{aligned}
 L_2 &= \frac{0.3 \times 0.191}{2} = 29 \text{ mH} \\
 L_3 &= \frac{1 - m^2}{4m} \left( \frac{Z_o}{3.14f_c} \right) \\
 &= \frac{1 - 0.75}{3.46} (0.19) = 0.014 \text{ H (or 14 mH)}
 \end{aligned} \tag{11.26}$$

$$\begin{aligned}
 C_2 &= \frac{m}{3.14Z_o f_c} \\
 &= \frac{0.866}{1.884 \times 10^6} = 0.459 \mu\text{F}
 \end{aligned} \tag{11.27}$$

The values of  $L_2$ ,  $L_3$ , and  $C_2$  are shown in Fig. 11.19*b*. The values of  $L_4$ ,  $L_5$ , and  $C_3$  will now be calculated.

$$\begin{aligned}
 L_4 &= \frac{mZ_o}{3.14f_c} \\
 &= \frac{0.6 \times 600}{3.14 \times 1000} = 0.114 \text{ H (or 114 mH)}
 \end{aligned} \tag{11.28}$$

$$\begin{aligned}
 L_5 &= \left( \frac{1 - m^2}{4m} \right) \left( \frac{Z_o}{3.14f_c} \right) \\
 &= \frac{(1 - 0.36) \times 0.190}{2.4} = 0.051 \text{ H (or 51 mH)}
 \end{aligned} \tag{11.26}$$

$$\begin{aligned}
 C_3 &= \frac{m}{3.14Z_o f_c} \\
 &= 0.6 \times 0.531 \times 10^{-6} = 0.319 \mu\text{F}
 \end{aligned} \tag{11.27}$$

The end sections are shown with the calculated values in Fig. 11.19*b*. The value of  $C_3$  is divided by two, and the value of  $L_5$  is multiplied by 2, as required by the specification shown in Fig. 11.19*a*.

In combining the four sections of Fig. 11.19*b*, the series inductances are added to make one single inductance value, thus reducing the number of components required. For example,  $L_4$  is added to  $L_1/2$  in combining the first and second sections. The final composite filter design is shown in Fig. 11.19*c*.

Although these filters are easy to design, they suffer from certain shortcomings. The most important is that the terminations must be such that at the center frequency of the filter the source and load must be pure resistances equal to the image impedance. As the frequency increases, the impedance must decrease to zero at the cutoff frequency. This cannot be done with lumped circuits, and so the actual band shape with resistive terminations is never exactly as shown on the graphs.

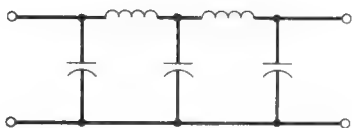
## 11.11 SOME ADDITIONAL TYPES OF FILTERS

Constant- $k$  and  $m$ -derived filters are easy to design (in comparison to other types), and they have reasonably good roll-off characteristics. Many other methods of designing LC filters are also possible, each with its own advantages and disadvantages.

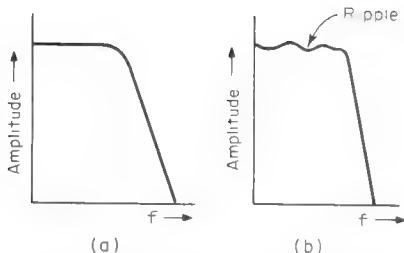
Filter manufacturers may make use of a computer to design a variety of filters having different attenuation and phase characteristics. The designers choose the filter with the characteristic most nearly equal to their required characteristics. This is comparable to the situation with tubes or transistors in which case the designer purchases the tube or transistor most nearly suited to his needs.

Since the computer has been made available to filter designers, there is a trend toward obtaining "tailor-made" filters. The designer specifies the characteristics he wants, and the computer performs the computation. What used to amount to hours and hours of tedious equation solving can now be accomplished in minutes. Also, more complex designs can now be obtained—designs that were not economically feasible when calculated without computers.

The designer of today, then, must know the characteristics, advantages, and disadvantages of the various types of filters. This enables him to specify filters so as to get the best possible design for his needs.



**Fig. 11.20** The configuration of a typical Butterworth or Chebyshev filter.



**Fig. 11.21** Typical attenuation response curves for Butterworth and Chebyshev filters: (a) Butterworth filter. (b) Chebyshev filter.

**BUTTERWORTH FILTER** This type of filter is also called a *maximally flat filter*. It is a form of *LC* filter that has a relatively flat response—that is, it offers a relatively constant impedance to signals with frequencies in its passband range.

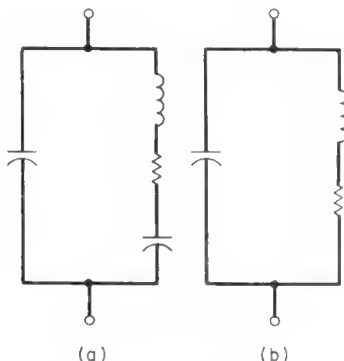
The configuration of a typical Butterworth filter is shown in Fig. 11.20. You will note that it is like a constant-*k* filter. However, the values of the components are different. This is to be expected because they are arrived at by a different method.

The Butterworth filter has a good phase response. This is a measure of the amount of phase shift that occurs when a signal passes through the filter. It also has good amplitude response, which is another way of saying that its response is flat. However, it has a rather gradual roll-off, which makes it unsuitable in applications where it is necessary to sharply cut off adjacent frequencies.

**CHEBYSHEV FILTER** The Chebyshev filter is also a form of *LC* filter in the same configuration as the Butterworth filter. The design of the Chebyshev filter is based on the assumption that all the frequencies to be passed are equally important. In this respect, it is different from the filters previously discussed which are designed on the assumption that zero frequency (dc) is the most important factor in design.

The cutoff characteristic of a Chebyshev filter is much sharper than that of the Butterworth or other types discussed (constant-*k* and *m*-derived). An important disadvantage of the Chebyshev is that there is a ripple in its passband response. This is another way of saying that its response curve is not flat in the passband region. Figure 11.21 compares the attenuation response curves of the Butterworth filter (which has no ripple) with that of the Chebyshev filter (which has ripple).

**CRYSTAL FILTER (SOMETIMES CALLED “QUARTZ FILTER”)** Figure 11.22 shows two equivalent circuits for crystals in their holders. The equivalent circuit of a piezo-

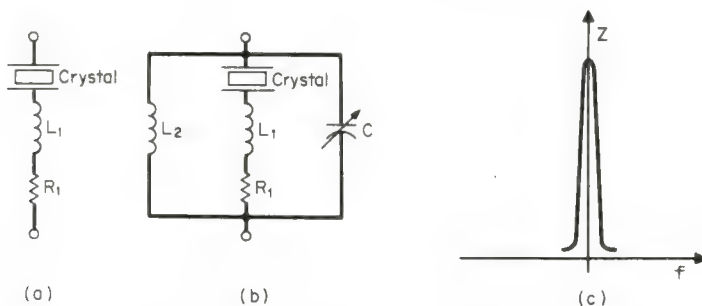


**Fig. 11.22** Equivalent circuits for a crystal: (a) This is the equivalent circuit that is usually shown. (b) Mathematically, this circuit can be obtained from the one shown in (a).

electric crystal that is usually shown is in Fig. 11.22a. The extremely high  $Q$  of a crystal causes it to behave as a bandpass filter with a very narrow bandwidth.

Mathematically, it is possible to convert the equivalent circuit of Fig. 11.22a into another equivalent circuit like the one shown in Fig. 11.22b. This one is easier to visualize, and calculations for this type of circuit were discussed in Chap. 10 of this book.

Figure 11.23 shows a crystal in filter circuit configurations. If a coil  $L_1$  with a resistance  $R_1$  is placed in series with the crystal, as shown in Fig. 11.23a, the circuit behaves as a simple parallel-resonant tuned circuit. It will pass all frequencies except the one to which it is tuned. Because of the high circuit  $Q$ , a resistor or coil may be placed in parallel with the circuit. This will lower the  $Q$  and broaden the frequency response. Also, a variable capacitor may be placed across the crystal circuit to tune it through a narrow range of frequencies. Figure 11.23b shows the practical circuit for a crystal filter, and Fig. 11.23c shows a typical response curve.



**Fig. 11.23** Crystal circuits used for filters: (a) A coil  $L_1$  and its dc resistance  $R_1$  in series with the crystal makes it behave as a simple parallel-resonant circuit with a high  $Q$ . (b) In this circuit the variable capacitor  $C$  tunes the circuit through a narrow range, and coil  $L_2$  lowers the circuit  $Q$  for a broader response.

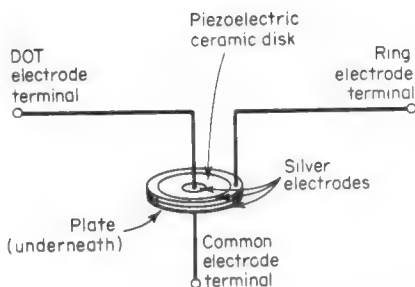
Crystal filters are used in applications where a parallel-resonant  $LC$  circuit response curve is applicable. For example, crystal filters are used in i-f amplifier coupling circuits to provide a narrow passband with sharp skirt selectivity. A number of crystals may be used in a stagger tuning arrangement to obtain a wider response than is normally obtainable with one crystal circuit. The very narrow bandwidth of a crystal filter is useful in CW receivers because their sharp selectivity makes possible separate signals of adjacent carrier frequencies.  $Q$  multipliers in the i-f stage may employ crystals for this purpose.

Disadvantages of crystal filters include their cost, which may be high compared to that of an  $LC$  filter. (However, if the  $LC$  filter was designed to produce exactly the same resonant curve as the crystal filter, it would be much bulkier and far more expensive.) Crystal filters cannot be tuned over a wide range of frequencies. They are bulkier than the ceramic filters that will be described next.

**CERAMIC FILTER** Ceramic filters have characteristics similar to those of quartz filters, but are much smaller in size. Instead of using a piece of quartz for the vibrating element, the ceramic filter uses the piezoelectric properties of lead zirconate-lead titanate resonators.

Figure 11.24 shows the construction of a simple three-terminal ceramic filter. The dot electrode terminal (in the center of the ceramic disk) is the one to which the input signal is delivered. The input signal causes the ceramic disk to vibrate at its mechanical resonant frequency. If the input signal frequency is other than the mechanical resonant frequency of the ceramic disk, no vibration will take place. When the ceramic disk is vibrating, voltages are generated around its edges which are picked off by the ring electrode terminal. The common terminal, which is simply a silver plate across the bottom of the ceramic disk, is normally operated at ground potential. As an indication of the very-small physical size of the ceramic filter, a ceramic disk 0.2 in. in diameter by 0.1 to 0.4 in thick will resonate at a frequency of 455 kHz.





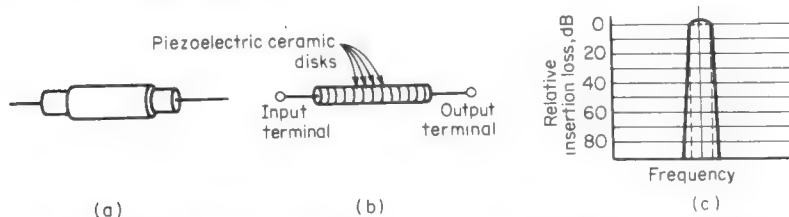
**Fig. 11.24** Construction details of a three-terminal ceramic filter.

A two-terminal ceramic filter is made by placing a silver plate on both sides of the piezoelectric ceramic disk. There is no common electrode terminal. In this type of filter, the signal is fed to one of the plates and taken from the other.

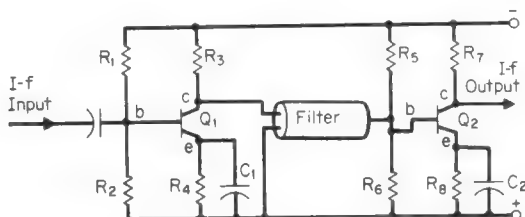
To get a sharper bandpass characteristic, a number of the two-terminal disks can be stacked to produce what is known as a *ceramic ladder filter*. Figure 11.25 presents this type of filter. Figure 11.25a shows its physical appearance, and Fig. 11.25b indicates how this type of filter is constructed by stacking ceramic wafers. Figure 11.25c gives the response curve of a ladder filter compared with that of an ideal filter (shown in dotted lines). Note that the shape of the ceramic ladder filter response curve is very near to the shape of the ideal curve.

Figure 11.26 demonstrates an i-f amplifier stage  $Q_1$  coupled to another i-f amplifier stage  $Q_2$  with ceramic filter coupling. Note that this is a three-terminal type filter, but it may also be a stacked filter. In other words, it is possible to stack three-terminal ceramic disks just as it is possible to stack the two-terminal type in order to get the desired bandpass characteristics.

The input signal to the filter in Fig. 11.26 is taken from collector-load resistor  $R_3$ , passed through the filter. The output signal is delivered to the base of  $Q_2$  at the junction of its base-bias voltage divider comprised of  $R_5$  and  $R_6$ . Although the circuit looks like a resistance-capacitance amplifier, remember that the filter is equivalent to an LC filter circuit.



**Fig. 11.25** The ladder-type ceramic filter and its characteristic curve. (a) Physical appearance of a ceramic ladder filter. (b) Construction of a ceramic ladder filter. (c) Response of a ceramic ladder filter.



**Fig. 11.26** An i-f stage that uses a ceramic filter for coupling.

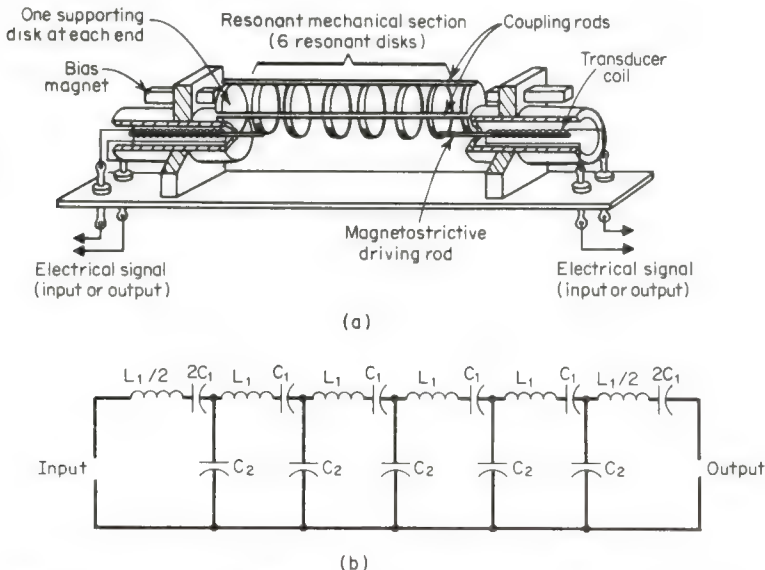
**MECHANICAL FILTERS (ALSO CALLED "MAGNETOSTRICTIVE FILTERS")** Both the quartz crystal and ceramic filters employ the piezoelectric effect to produce a mechanical vibration in the filter material. This mechanical vibration, which determines the frequency of the filter, is directly related to its physical size and can be held to a very close frequency tolerance.

Instead of the combination of piezoelectric principle and mechanical resonance, it is also possible to use the magnetostrictive principle. The two magnetostrictive effects that are related to the use of mechanical filters are the *Joule effect* and the *Vallari effect*. The Joule effect relates to the change in physical length of a ferromagnetic material in the presence of an applied magnetic field. The Vallari effect relates to the change in magnetization of a material when an external stress is applied to that material. Both these effects are usually combined and referred to as the magnetostrictive effect.

Figure 11.27 presents a mechanical filter and its equivalent  $LC$  circuit. The construction details of a mechanical filter are shown in Fig. 11.27a. The electric input signal produces a varying magnetic field around a magnetostrictive driving rod. This varying field causes the driving rod to change in length at a rate that is directly equal to the rate at which the field varies. The varying length of the rod produces a mechanical vibration in the resonant mechanical section. Note that the resonant section is comprised of metal disks which are manufactured to be precisely resonant at a certain frequency. These disks are coupled together by rigid coupling rods.

Any frequency other than the resonant frequency cannot pass through the mechanical filter because the resonant frequency of the disks is quite precise—that is, they are manufactured to very close physical dimensions. Thus, the mechanical filter cannot be set into vibration by frequencies other than the mechanical frequencies. The mechanical oscillations are converted back to electric oscillations by a transducer output coil.

The bias magnets at each end produce a restraining force on the motion of the disks in order to make the filter vibrate only once for each input cycle. Remember that one complete cycle of ac signal consists of two peak currents. Each of these peaks will magnetize the rod and cause it to change its length. Thus, two vibrations would occur for each cycle if it were not for the biasing magnet. The magnetic field from the bias magnet holds the resonant disks so that they can move easily in one direction, but cannot move easily in the opposite direction. With this arrangement the output frequency is the same as the input frequency.



**Fig. 11.27** Construction details of a mechanical filter and its equivalent circuit: (a) Elements of a mechanical filter. (b) Electrical analogy of a mechanical filter.

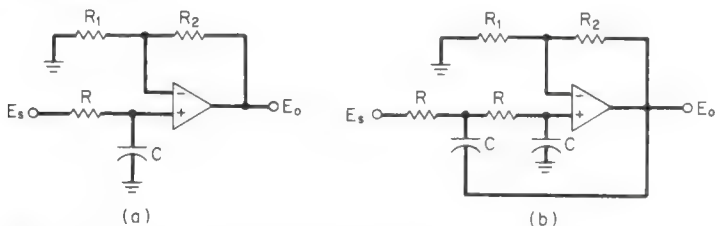
The electric equivalent circuit in Fig. 11.27*b* shows that the mechanical filter is comparable to a ladder-type  $LC$  filter.

A variation in design of the mechanical filter, which is used at much lower frequencies, employs a tuning fork instead of vibrating mechanical metal disks. The input signal is electromagnetically coupled to one prong of the tuning fork which sets it into vibration. The output signal is taken from the other prong. Mechanical filters of this type are used at audio frequencies.

### 11.12 ACTIVE FILTERS

As mentioned in the beginning of the chapter, an active filter contains, in addition to op amps, resistors and capacitors. Inductors, a component in  $LC$  filters, are thereby eliminated. Because of their relative large size and cost, the elimination of inductors in filters is indeed a considerable achievement. This is especially true for filters having cutoff frequencies of 100 Hz or less.

There are many methods one may use for the realization of active filters. In this section we shall concentrate on an active filter that provides the Butterworth response of Fig. 11.21. Two basic low-pass circuits for realizing Butterworth active filters are illustrated in Fig. 11.28.



**Fig. 11.28** Basic low-pass Butterworth filters: (a) First-order section. (b) Second-order section.

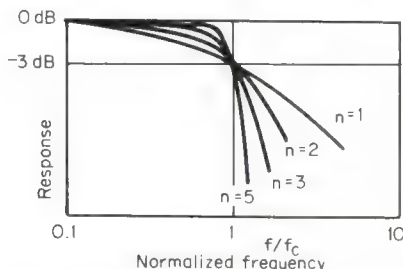
A first-order low-pass section is shown in Fig. 11.28*a* and a second-order low-pass section in Fig. 11.28*b*. The voltage gain,  $A_v$ , of each section, from Chap. 13, is determined by resistors  $R_1$  and  $R_2$  connected to the inverting ( $-$ ) terminal of the op amp:

$$A_v = 1 + \frac{R_2}{R_1} \quad (11.37)$$

The half-power cutoff frequency  $f_c$  for each section is determined from the following expression,

$$f_c = \frac{0.159}{RC} \quad (11.38)$$

where  $f_c$  is in hertz,  $R$  in ohms, and  $C$  in farads. The basic sections of Fig. 11.28, as



**Fig. 11.29** Frequency-response curves of a Butterworth low-pass filter for different orders,  $n$ .

will be explained shortly, may also be used to realize high-pass, bandpass, and band-reject filters.

Figure 11.29 provides plots of frequency response of a low-pass Butterworth filter for different orders,  $n$ . A particular value of  $n$  is obtained by cascading the basic sections of Fig. 11.28. As  $n$  increases, the shape factor improves. The frequency scale is normalized by plotting the ratio of  $ff_c$ . At  $f = f_c$  ( $ff_c = 1$ ), the response of the filter is down by  $-3$  dB.

For a first-order ( $n = 1$ ) Butterworth low-pass filter, the first-order section of Fig. 11.28a is used. For a second-order ( $n = 2$ ) filter, the second-order section of Fig. 11.28b is employed. If a third-order ( $n = 3$ ) filter is necessary, the first-order and second-order sections of Fig. 11.28 are cascaded. A fourth-order ( $n = 4$ ) filter is realized by cascading two second-order sections, and so on.

The gain of the first-order section is arbitrary. The gain of a second-order section, however, is defined. It is related by coefficient  $a$ , listed in Table 11.1, for different values of  $n$ . The relationship between  $A_v$  and  $a$  is

$$A_v = 3 - a \quad (11.39)$$

Example 11.6 illustrates the use of the preceding equations and Table 11.1 in a filter design problem.

**TABLE 11.1 Coefficients for a Second-Order Butterworth Section**

$n$	$a$
2	1.414
3	1
4	0.765, 1.848
5	0.618, 1.618
6	0.518, 1.414, 1.932

patch p. 11-27

**example 11.6** Design a fifth-order ( $n = 5$ ) Butterworth low-pass filter having a cutoff frequency of 1 kHz. For each section,  $R = 1$  k $\Omega$  and  $R_1 = 5$  k $\Omega$ . The first-order section has a gain of 10.

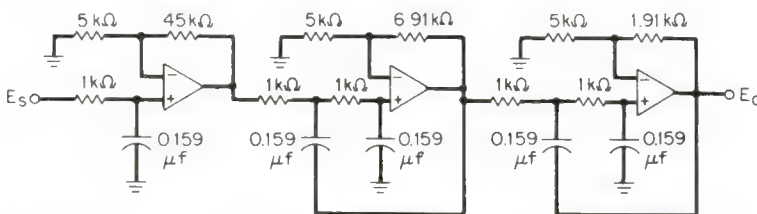
**solution** The filter consists of a cascaded connection of a first-order and two second-order sections, as illustrated in Fig. 11.30. By Eq. (11.28), for  $R = 1$  k $\Omega$ ,

$$\begin{aligned} C &= \frac{0.159}{f_c R} \\ &= \frac{0.159}{10^3 \times 10^3} = 0.159 \times 10^{-6} \text{ F} = 0.159 \text{ } \mu\text{F} \end{aligned}$$

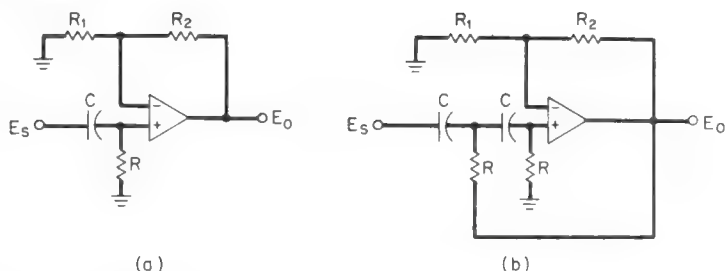
For a gain of 10 for the first-order section, by Eq. (11.37),

$$10 = 1 + \frac{R_2}{5 \text{ k}}$$

Solving,  $R_2 = 45$  k $\Omega$ .



**Fig. 11.30** An example of a fifth-order ( $n = 5$ ) 1-kHz low-pass Butterworth active filter. (See Example 11.6.)



**Fig. 11.31** Basic high-pass Butterworth filters: (a) First-order section. (b) Second-order section.

Referring to Table 11.1, for the first second-order section in Fig. 11.30,  $a = 0.618$ . By Eq. (11.39),

$$A_{v1} = 3 - 0.618 = 2.382$$

By Eq. (11.37) then,

$$2.382 = 1 + \frac{R_2}{5 \text{ k}}$$

and  $R_2 = 6.91 \text{ k}\Omega$ .

For the remaining second-order section, from Table 11.1,  $a = 1.618$ . Hence,

$$A_{v2} = 3 - 1.618 = 1.382$$

$$1.382 = 1 + \frac{R_2}{5 \text{ k}}$$

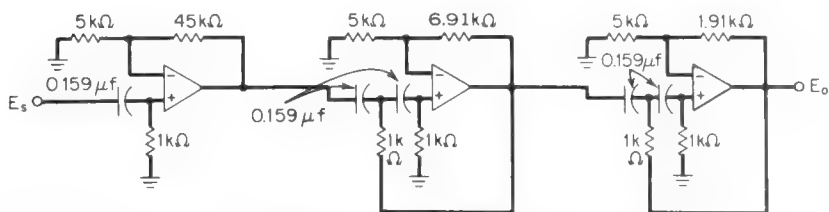
Solving,  $R_2 = 1.91 \text{ k}\Omega$ .

If, for example, a sixth-order ( $n = 6$ ) filter were required, three second-order sections of Fig. 11.28b would be cascaded. In calculating the values of  $R_2$ , from Table 11.1,  $a = 0.518$  would be used for the first section, 1.414 for the second, and 1.932 for the last section.

**High-pass filter** If the  $R$ 's and  $C$ 's in Fig. 11.28 are interchanged, first-order and second-order high-pass sections are obtained. These are illustrated in Fig. 11.31. Table 11.1 and Eqs. (11.37) to (11.39) also apply to the high-pass sections. This is illustrated in the following example.

**example 11.7** Design a fifth-order ( $n = 5$ ) Butterworth high-pass filter having a cutoff frequency of 1 kHz. For each section,  $R = 1 \text{ k}\Omega$  and  $R_1 = 5 \text{ k}\Omega$ . The first stage has a voltage gain of 10.

**solution** The filter, shown in Fig. 11.32, uses the basic high-pass sections of Fig. 11.31. The values of  $C$  and  $R_2$  were obtained in the manner illustrated in Example 11.6 for the fifth-order low-pass filter.



**Fig. 11.32** An example of a fifth-order 1-kHz high-pass Butterworth active filter. (See Example 11.7.)

**Bandpass filter** An active bandpass filter may be obtained by cascading high- and low-pass active filters, as illustrated in the block diagram of Fig. 11.33. The cutoff frequency of the low-pass filter must be greater than the cutoff frequency of the high-pass filter. If the cutoff frequencies for the high-pass and low-pass filters are  $f_H$  and  $f_L$ , respectively, the bandwidth of the bandpass filter is  $f_L - f_H$ .



Fig. 11.33 Block diagram of a bandpass active filter.

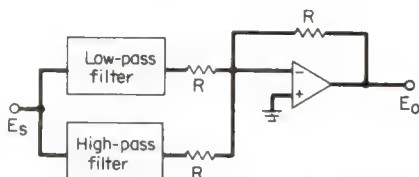


Fig. 11.34 Realizing an active band-reject filter by connecting a low-pass and a high-pass filter to a summer.

**Band-reject filter** An active band-reject filter may be realized by connecting together the inputs of a low-pass and a high-pass filter and their respective outputs to a summing amplifier. This is illustrated in Fig. 11.34. For this filter,  $f_L$  must be less than  $f_H$ . The stop band is then equal to  $f_H - f_L$ .





# Chapter 12

## Transistor Amplifiers and Oscillators

### 12.1 INTRODUCTION

An amplifier increases the magnitude of, or *amplifies*, an electric signal. The signal may be derived, for example, from a TV antenna, a phono cartridge, or a strain gage. Of all components, the amplifier is by far the most widely used building block in electronics systems.

An oscillator generates a waveform without the presence of an external signal. If the waveform is a sine wave, the oscillator is referred to as a *sinusoidal* oscillator. For a waveform other than a sine wave, such as a square or sawtooth wave, the oscillator is called a *nonsinusoidal*, or *relaxation*, oscillator. Sinusoidal oscillators are used, for example, to test the performance of circuits, to generate radio frequencies, and in receivers.

In this chapter we shall consider how amplifiers are classified, their characteristics, biasing, how to calculate their gain and frequency response, and the effects of utilizing more than one amplifier stage. The various parameters used for characterizing transistors are defined in Chap. 8. Feedback, where a portion or all of the output signal is returned to the input of an amplifier, and the operation of sinusoidal oscillators, conclude the chapter.

### 12.2 CLASSIFICATION OF AMPLIFIERS

An amplifier may be represented by the simple block diagram of Fig. 12.1. The input signal voltage is denoted by  $E_s$  and the input signal current by  $I_s$ . Across the output terminals is load resistor  $R_L$ . Resistance  $R_L$  may represent, for example, the resistance of a loudspeaker coil, a motor winding, or the input of a connected amplifier stage. Voltage  $E_o$  is the output (load) voltage, and current  $I_o$  the output (load) current. Symbols  $P_i$  and  $P_o$  are the input and output signal powers, respectively.

If the amplifier is optimized to amplify voltage signals, it is called a *voltage* amplifier. For current signals, it is referred to as a *current* amplifier. If it is to develop output power of one watt or more, it is generally classified as a *power* amplifier.

Another method of classification is based on the amplitude of signals. For small signals, in the order typically of microvolts and millivolts, the amplifier may be classified as *small-signal*, and for larger signals as a *large-signal* amplifier. Simple algebra is all that is required for calculating the performance of small-signal amplifiers. Graphical methods, however, are generally used in the analysis of large-signal amplifiers.

Amplifiers are also classified in terms of load current flow. Figure 12.2a shows one cycle (360°) of a sine wave input signal to an amplifier. If the load current also flows

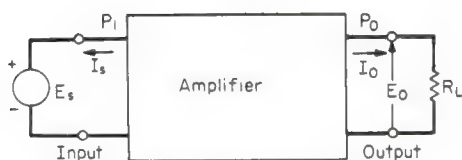


Fig. 12.1 Block diagram of an amplifier.

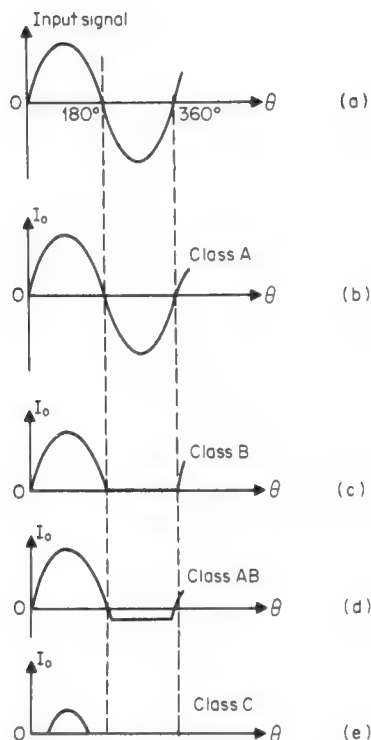


Fig. 12.2 Classes of amplifier operation: (a) Input signal. (b) Load current flows  $360^\circ$  for class A operation. (c) Load current flows  $180^\circ$  for class B operation. (d) In class AB operation, load current flow is greater than  $180^\circ$  but less than  $360^\circ$ . (e) Class C operation, where load current flows for less than  $180^\circ$ .

for  $360^\circ$ , as indicated in Fig. 12.2b, the amplifier is operating as *class A*. *Class B* operation, illustrated in Fig. 12.2c, exists when the load current flows for only  $180^\circ$ . If the load current flows for more than  $180$  but less than  $360^\circ$ , *class AB* operation is realized (Fig. 12.2d). *Class C* operation is obtained if the load current flows for less than  $180^\circ$  (Fig. 12.2e).

Class A operation is used for small-signal and single-transistor (referred to as *single-ended*) power amplifiers. If class B, AB, or C operation is used in a single-ended amplifier, the output waveform is badly distorted and the amplifier is worthless. For this reason, for class B or AB operation the amplifier contains two transistors, referred to as a *push-pull* amplifier. Class C operation is often used in the amplification of r-f signals in a circuit containing a tuned load.

Amplifiers are also classified according to their intended operation. Examples include *audio*, *r-f*, *video*, *microwave*, and *pulse* amplifiers.

### 12.3 CHARACTERISTICS OF AMPLIFIERS

The performance of amplifiers is characterized by the following terms:

- Voltage gain
- Current gain
- Power gain
- Input resistance

Output resistance  
 Bandwidth  
 Distortion  
 Slewing rate

**VOLTAGE GAIN** The voltage gain of an amplifier  $A_v$  is defined as the ratio of the output signal voltage  $E_o$  to the input signal voltage  $E_s$ :

$$A_v = \frac{E_o}{E_s} \quad (12.1)$$

**CURRENT GAIN** The current gain of an amplifier  $A_i$  is defined as the ratio of the output signal current  $I_o$  to the input signal current  $I_s$ :

$$A_i = \frac{I_o}{I_s} \quad (12.2)$$

**POWER GAIN** The power gain of an amplifier  $A_p$  is defined as the ratio of the output signal power  $P_o$  to the input signal power  $P_i$ :

$$A_p = \frac{P_o}{P_i} \quad (12.3a)$$

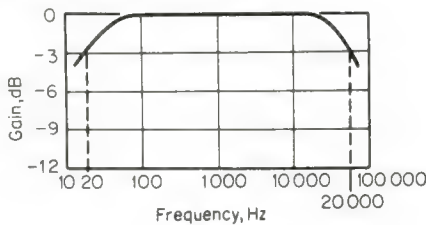
An alternative definition is that the power gain is equal to the product of the voltage and current gains:

$$A_p = A_v A_i \quad (12.3b)$$

**INPUT RESISTANCE** The input resistance (or impedance) of an amplifier  $R_i$  (or  $Z_i$ ) is the resistance (or impedance) across the input terminals of an amplifier.

**OUTPUT RESISTANCE** The output resistance (or impedance) of an amplifier  $R_o$  (or  $Z_o$ ) is the resistance (or impedance) across the output terminals of an amplifier with the signal source set to zero (see Chap. 6).

**BANDWIDTH** The bandwidth of an amplifier BW is a band of frequencies that is amplified with modest attenuation in gain. A typical bandwidth curve is illustrated in Fig. 12.3. Gain, expressed in decibels (dB), is generally plotted along the  $y$  axis, and frequency along the  $x$  axis. To cover conveniently a large range of frequencies, a logarithmic scale is used for the frequency axis.



**Fig. 12.3** An example of bandwidth characteristics of an amplifier.

The nominal gain over a large range of frequencies is expressed as 0 dB. At frequencies  $f_L = 20$  Hz and  $f_H = 20$  kHz, the gain is down by -3 dB. Therefore, frequency  $f_L$  is referred to as the *lower* -3-dB frequency, and  $f_H$  as the *upper* -3-dB frequency. Other terms used include the *lower* and *upper break*, *cutoff*, *corner*, and *half-power* frequencies. The bandwidth is equal to the difference between the upper and lower -3-dB frequencies:

$$BW = f_H - f_L \quad (12.4)$$

From Fig. 12.3,  $BW = 20\,000 - 20 \approx 20\,000$  Hz = 20 kHz.

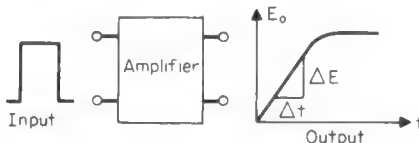
## 12-4 Transistor Amplifiers and Oscillators

**DISTORTION** Because the characteristics of the BJT and FET are not strictly linear (see Chap. 8), the output of an amplifier contains frequencies that were not present in the input signal. Assume, for example, that a 1-kHz sine wave is impressed across the input terminals of an amplifier. The output will contain, in addition to a 1-kHz frequency, referred to as the *fundamental frequency*, integrally related frequencies of 2, 3, 4, . . . , kHz. These frequencies are referred to as *harmonics*: 2 kHz is the *second harmonic*, 3 kHz is the *third harmonic*, and so on.

**Harmonic distortion** Harmonic distortion is defined as the ratio of the amplitude of a harmonic component to the amplitude of the fundamental frequency. For example, if the 1-kHz output is 10 V and the 2-kHz output is 0.5 V, the *second-harmonic distortion*  $D_2$  is  $0.5/10 = 0.05 = 5$  percent.

**Intermodulation distortion** Assume that a 1000-Hz and a 400-Hz signal are fed into an amplifier. The output, in addition to the 1000- and 400-Hz signals, will contain frequencies equal to their *difference* ( $1000 - 400 = 600$  Hz) and to their *sum* ( $1000 + 400 = 1400$  Hz). The sum and difference frequencies is the result of intermodulation distortion in an amplifier.

**SLEWING RATE** The slewing rate of an amplifier indicates how well it responds to a rapidly changing waveform. To measure the response, a rectangular pulse is applied to the input of the amplifier (Fig. 12.4). Because of internal capacitances of



**Fig. 12.4** Determining the slewing rate of an amplifier. The value of the slewing rate equals  $\Delta E/\Delta t$ .

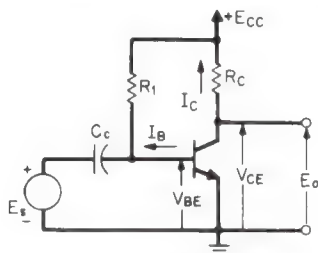
the amplifier, it takes a finite time for the output to reach its maximum value. How fast it reaches it is the slewing rate of the amplifier.

As indicated in Fig. 12.4, an increment of time  $\Delta t$  ( $\Delta$  is the Greek letter *delta*) is required for the output voltage to increase by the increment  $\Delta E$  along its linear ascent. The ratio  $\Delta E/\Delta t$  is equal to the slewing rate. Its unit is generally expressed in volts per microsecond.

### 12.4 BIASING AND STABILIZATION

Because of its wide use, in this section we consider how the BJT and FET amplifiers are biased and stabilized for class A operation.

**BIASING THE COMMON-EMITTER AMPLIFIER** A typical common-emitter (CE) junction transistor amplifier is illustrated in Fig. 12.5. A single bias source,  $E_{CC}$ , sup-



**Fig. 12.5** Basic class A common-emitter amplifier biased from a single dc source,  $E_{CC}$ .

plies the necessary collector and base currents for class A operation. Capacitor  $C_c$ , called a *coupling capacitor*, blocks any direct current from the base that may be in signal source  $E_s$ . Output signal voltage  $E_o$  is taken across the collector and ground.

A simple method of finding the *quiescent operating point Q* is by graphical analysis. First, summing the dc voltages in the collector circuit of Fig. 12.5, we have

$$E_{CC} = R_C I_C + V_{CE} \quad (12.5a)$$

Solving Eq. (12.5a) for  $I_C$ ,

$$I_C = -\frac{V_{CE}}{R_C} + \frac{E_{CC}}{R_C} \quad (12.5b)$$

Expression (12.5b) is an equation of a straight line of the form  $y = mx + b$ , where  $m$  is the slope and  $b$  is the  $y$  intercept. Comparing the equation of a straight line with Eq. (12.5b), we see that  $y = I_C$ ,  $m = -1/R_C$ ,  $x = V_{CE}$ , and  $b = E_{CC}/R_C$ . Because only two points determine a straight line, it is convenient to superimpose Eq. (12.5b) on the collector characteristics of a transistor. The superimposed line is referred to as a *load line*.

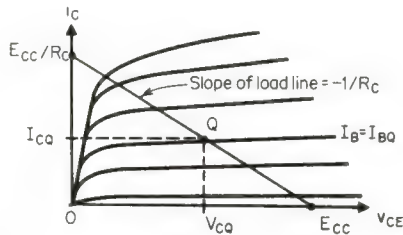
To obtain the two points, Eq. (12.5b) is first solved for  $I_C = 0$ ; then

$$V_{CE} = E_{CC} \quad (12.6a)$$

This is one point needed to plot the load line. The second point is obtained by setting  $V_{CE} = 0$ . For this condition, Eq. (12.5b) reduces to

$$I_C = \frac{E_{CC}}{R_C} \quad (12.6b)$$

The two points and the load line drawn between them on the collector characteristics are illustrated in Fig. 12.6. The slope of the load line is  $-1/R_C$ . Location of the  $Q$



**Fig. 12.6** Superimposing load line on collector characteristics for location of the  $Q$  point.

point is at the intersection of the load line and base current curve  $I_{BQ}$ . For class-A operation the  $Q$  point is located to be approximately at the center of the load line. The quiescent collector voltage and current are denoted by  $V_{CQ}$  and  $I_{CQ}$ , respectively.

Returning to Fig. 12.5, the equation for the dc quantities in the input circuit is

$$E_{CC} = I_B R_1 + V_{BE} \quad (12.7a)$$

where, at room temperature,  $V_{BE} = 0.7$  V for a silicon transistor and 0.3 V for a germanium transistor (see Chap. 8).

Because the bipolar junction transistor is a *current-operated* device, it is biased by a base current  $I_B$ . Solving Eq. (12.7a) for  $I_B$ , we obtain

$$I_B = \left( \frac{E_{CC} - V_{BE}}{R_1} \right) \quad (12.7b)$$

The application of this and the other expressions derived in the section is illustrated in Example 12.1.



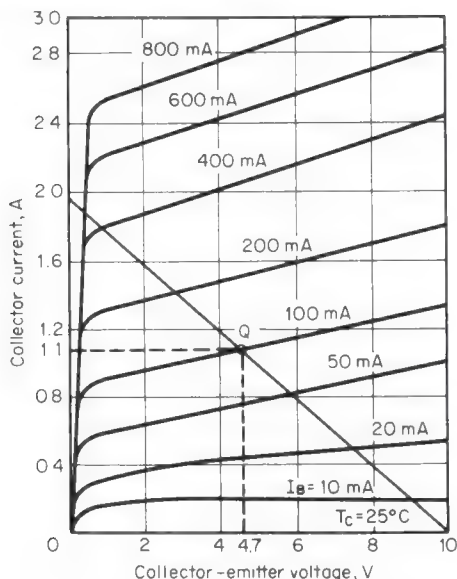


Fig. 12.7 Locating the Q point in Example 12.1.

**example 12.1** In the circuit of Fig. 12.5,  $E_{CC} = 10$  V,  $R_C = 5 \Omega$ ,  $I_B = 100$  mA, and  $V_{BE} = 0.7$  V. Assume that the collector characteristics of Fig. 12.7 are appropriate for the transistor. (a) Write the equation for the load line. (b) Superimpose the load line on the collector characteristics. (c) Locate the Q point. (d) Determine the value of  $R_1$ .

**solution** (a) Substitution of  $R_C = 5 \Omega$  and  $E_{CC} = 10$  V in Eq. (12.5b) yields  $I_C = -V_{CE}/5 + 10/5 = -V_{CE}/5 + 2$ , which is the equation of the load line.

(b) For  $I_C = 0$ , by Eq. (12.6a),  $V_{CE} = 10$  V. For the second point, letting  $V_{CE} = 0$ , by (12.6b),  $I_C = 10/5 = 2$  A. The load line drawn between the two points is shown in Fig. 12.7.

(c) From Fig. 12.7, the Q point is located at the intersection of the load line and the  $I_B = 100$  mA curve:  $V_{CEQ} \approx 4.7$  V and  $I_{CQ} \approx 1.1$  A.

(d) Solving Eq. (12.7b) for  $R_1$ ,

$$R_1 = \frac{E_{CC} - V_{BE}}{I_B}$$

Substitution of the given values in the equation yields

$$R_1 = \frac{10 - 0.7}{0.1} = 93 \Omega$$

If collector characteristics are not available, the Q point can be estimated. The procedure is illustrated in the next example.

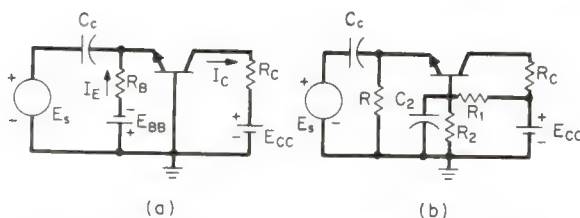
**example 12.2** Using the values given in Example 12.1 and the fact that the dc current gain,  $h_{FE} = 10$ , find the Q point of the amplifier of Fig. 12.5.

**solution** Collector current  $I_C$  equals the product of base current  $I_B$  and the dc current gain  $h_{FE}$ :  $I_C = I_B h_{FE}$ . From Example 12.1,  $I_B = 100$  mA = 0.1 A; hence,  $I_C = I_{CQ} = 0.1 \times 10 = 1$  A. Solving Eq. (12.5a) for  $V_{CE}$ , we obtain

$$V_{CE} = E_{CC} - R_C I_C$$

Therefore,  $V_{CE} = V_{CEQ} = 10 - 5 \times 1 = 5$  V. The Q point is located at  $V_{CEQ} = 5$  V and  $I_C = 1$  A. These values are reasonably close to those obtained graphically in Example 12.1.

**BIASING THE COMMON-BASE AMPLIFIER** Generally, two dc sources are used for biasing the common-base (CB) amplifier. Required are  $E_{CC}$  for the output circuit and  $E_{BB}$  for the input circuit, as shown in Fig. 12.8a. Because the current gain of a transistor in the common-base configuration is close to unity, the common-base collector

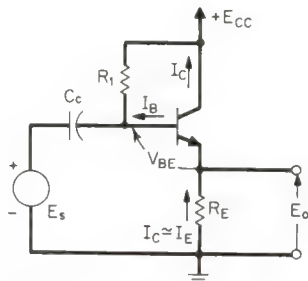


**Fig. 12.8** Biasing a common-base amplifier: (a) Using two dc sources,  $E_{CC}$  and  $E_{BB}$ . (b) Using a single source,  $E_{CC}$ .

characteristics are not very informative. The  $Q$  point is therefore calculated in a manner similar to that illustrated in Example 12.2. The dc current gain  $h_{FB}$  of a CB amplifier is taken to be equal to one:  $h_{FB} = 1$ .

A method for biasing a common-base amplifier from a single dc source  $E_{CC}$  is shown in Fig. 12.8b. Bypass capacitor  $C_2$  is chosen so that its reactance to the lowest signal frequency is at least one-tenth the value of the resistance of  $R_2$ . This ensures that the base is common to the input and output signals. For direct current, however, the capacitor acts like an "open circuit," and the transistor is biased as a common-emitter amplifier. Resistor  $R$  provides a dc return path between the emitter and ground.

**BIASING THE EMITTER FOLLOWER** A basic emitter follower, also referred to as a *common-collector* (CC) amplifier, is illustrated in Fig. 12.9. The collector is connected directly to dc bias source  $E_{CC}$ . Output signal voltage  $E_o$  is taken across the emitter resistor  $R_E$ . Because the emitter and collector currents are approximately the same in a transistor, the common-emitter collector characteristics can be used in the analysis of the emitter follower.



**Fig. 12.9** Basic emitter follower circuit.

Writing an equation for the dc voltages around the input circuit, we have

$$E_{CC} = R_1 I_B + V_{BE} + R_E I_C$$

assuming that  $I_C$  is approximately equal to the emitter current  $I_E$ . Solving for  $I_C$ ,

$$I_C = \frac{E_{CC} - R_1 I_B - V_{BE}}{R_E} \quad (12.8)$$

The dc voltage across the emitter resistor  $E_{o(dc)}$  is

$$E_{o(dc)} = R_E I_C \quad (12.9)$$

**STABILIZING THE COMMON-EMITTER AMPLIFIER** Because of variations in reverse saturation current, current gain, and the base-emitter voltage with temperature and aging, it is necessary to *stabilize* the  $Q$  point of most common-emitter amplifiers. (The common-base and emitter follower amplifiers may be considered to be inherently stable.) A number of stabilization techniques exist and are examined in this section.

## 12-8 Transistor Amplifiers and Oscillators

**Current-feedback stabilization** An example of current-feedback stabilization is illustrated in Fig. 12.10a. Resistors  $R_1$  and  $R_2$  establish the  $Q$  point. Emitter resistor  $R_E$ , connected between the emitter of the transistor and ground, provides stabilization. In analyzing the circuit, it is convenient to replace resistors  $R_1$  and  $R_2$  connected to  $E_{CC}$  with a Thévenin equivalent circuit (see Chap. 6). Thévenin resistance  $R_B$  is

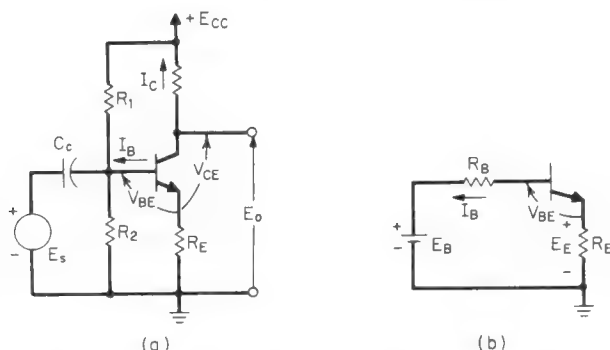
$$R_B = R_1 || R_2 = \frac{R_1 R_2}{R_1 + R_2} \quad (12.10a)$$

The Thévenin voltage  $E_B$  is

$$E_B = \frac{E_{CC} R_2}{R_1 + R_2} \quad (12.10b)$$

The Thévenin circuit is given in Fig. 12.10b.

Assume that the dc current gain increases beyond its initial value. As a result, the collector current tends to rise, and the voltage across the emitter resistor  $E_E$  increases.



**Fig. 12.10** Common-emitter amplifier with current-feedback stabilization: (a) Circuit. (b) Biasing network replaced by its Thévenin equivalent circuit.

Because voltage  $E_E$  opposes  $E_B$ , the base-emitter voltage tends to decrease. Less base current flows, and the collector current is reduced to approximately its initial value.

If, for some reason, the collector current tends to decrease, voltage  $E_E$  also decreases. The base-emitter voltage increases, and more base (and collector) current flows.

From the previous description, current-feedback stabilization provides “automatic regulation” of the location of the  $Q$  point. A useful criterion for the effectiveness of current-feedback stabilization is the *current stability factor*  $S_I$ . It relates changes in collector current to changes in the reverse saturation current. A conservative approximate expression for  $S_I$  is

$$S_I \approx 1 + \frac{R_B}{R_E} \quad (12.11)$$

where, referring to Fig. 12.10b,  $R_B$  is the resistance of the parallel combination of  $R_1$  and  $R_2$  and  $R_E$  is the emitter resistance.

It is noted that by Eq. (12.11), if  $R_B$  is zero or  $R_E$  is infinite,  $S_I = 1$ , which represents perfect stabilization. For either of these values, however, the signal gain would be zero. Practical values for  $S_I$  are in the range of 4 to 10.

**example 12.3** If, in Fig. 12.10a,  $R_1 = 60$ ,  $R_2 = 12$ , and  $R_E = 2$  k $\Omega$ , calculate the value of  $S_I$ .

**solution** By Eq. (12.10a),  $R_B = R_1 || R_2 = (60 \times 12)/(60 + 12) = 10$  k $\Omega$ . By Eq. (12.11),  $S_I = 1 + \frac{10}{2} = 1 + 5 = 6$ .

**Voltage-feedback stabilization** Another stabilization technique, voltage-feedback stabilization, is illustrated in Fig. 12.11. The feedback resistor  $R_F$ , connected between the collector and base of the transistor is the key element.

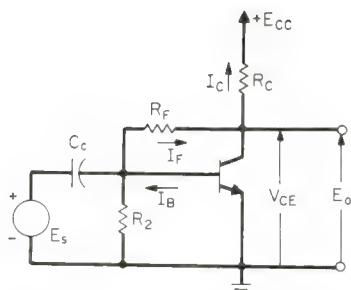


Fig. 12.11 Common-emitter amplifier with voltage-feedback stabilization.

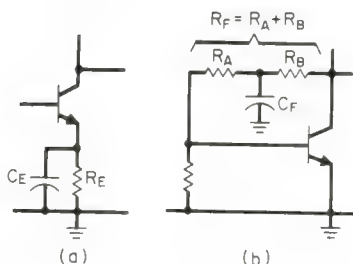


Fig. 12.12 Offsetting the reduction in signal gain of an amplifier with current- or voltage-feedback stabilization: (a) Bypassing emitter resistance. (b) Decoupling feedback network.

Assume that, because of high-temperature operation, the base-emitter voltage has been reduced. Base current and, as a result, collector current increase. Consequently, the collector-emitter voltage  $V_{CE}$  is reduced. Because the feedback current  $I_F$  in the feedback resistor is approximately equal to the collector-emitter voltage divided by the feedback resistance ( $I_F \approx V_{CE}/R_F$ ), the feedback current falls. This, in turn, causes the base and collector currents to decrease. The  $Q$  point, therefore, tends to return to its initial value. Changes in reverse saturation current or dc current gain, whether increasing or decreasing, are stabilized in a similar manner.

The current stability factor for voltage-feedback stabilization is given by

$$S_I \approx 1 + \frac{R_F}{R_C} \quad (12.12)$$

For good stability  $R_F$  should be low and  $R_C$  high in value. A low value of feedback resistance, however, reduces the signal gain of the amplifier.

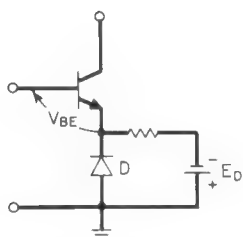
**Bypass and decoupling capacitors** To realize a low value of stability factor without appreciable deterioration in signal gain, *bypass* and *decoupling* capacitors are employed. In Fig. 12.12a, bypass capacitor  $C_E$  is in parallel with emitter resistor  $R_E$ . If the reactance of  $C_E$  is one-tenth or less of the value of  $R_E$  at the lowest signal frequency, the signal gain of the amplifier is hardly affected.

In Fig. 12.12b, resistor  $R_F$  is replaced by two resistors,  $R_A$  and  $R_B$ . Their sum is made equal to  $R_F$  ( $R_F = R_A + R_B$ ). Often,  $R_A$  and  $R_B$  are chosen to be equal to one-half the value of  $R_F$ . Decoupling capacitor  $C_F$  is connected between the junction of  $R_A$ ,  $R_B$  and ground. For minimum effect on signal gain,  $C_F$  is chosen so its reactance is one-tenth the value of  $R_A$  or  $R_B$ , whichever is smaller, at the lowest signal frequency.

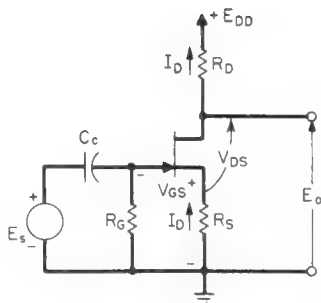
**Diode stabilization** A method employing a junction diode for stabilizing a common-emitter amplifier is illustrated in Fig. 12.13. The cathode of junction diode  $D$  is connected to the emitter of the transistor. Source  $E_D$  forward-biases the diode. It is reasonable to expect that variation in the diode junction voltage will track and oppose variations in the base-emitter voltage of the transistor. Consequently, changes in base-emitter voltage are canceled. In a silicon transistor, the most commonly used BJT, changes in the  $Q$  point are very dependent on the base-emitter voltage.

**BIASING JFET AND DEPLETION-TYPE MOSFET AMPLIFIERS** Because the FET is a voltage-operated device, *self-biasing* may be used. This is the prime method for biasing the JFET or depletion-type MOSFET. An example of an n-channel FET, self-biased by source resistor  $R_S$  connected between the source terminal of the FET and ground, is shown in Fig. 12.14. Drain current  $I_D$  results in a voltage drop across  $R_S$  which is equal to  $I_D R_S$ . The polarity of the voltage drop is such that the source is positive with respect to the ground.

Resistor  $R_G$  (typical value is 100 k $\Omega$ ) is connected between the gate terminal and ground. It provides a dc return path for the voltage across the source resistor. As a



**Fig. 12.13** Example of diode stabilization for a common-emitter amplifier.



**Fig. 12.14** An n-channel common-source FET amplifier which is self-biased by source resistor  $R_S$ .

result, the gate is biased *negatively* with respect to the source by  $-I_D R_S$  volts. For a p-channel device, the polarity of  $E_{DD}$  is reversed and the gate-source voltage is  $+I_D R_S$  volts.

In addition to providing a bias voltage, the source resistor also stabilizes the operation of the amplifier. Its behavior is analogous to the emitter resistor used in current feedback stabilization of a BJT amplifier. To minimize the decrease in signal gain,  $R_S$  is bypassed with a capacitor, as in the common-emitter amplifier.

Writing an equation for the dc voltages in the output side of Fig. 12.14, we obtain

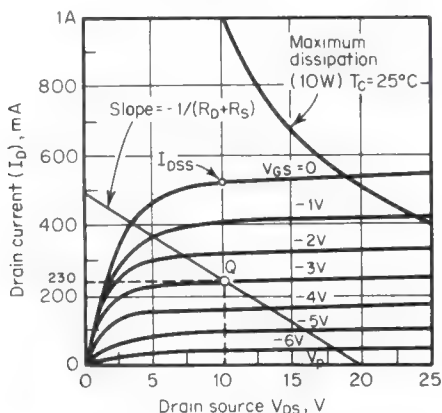
$$E_{DD} = I_D(R_D + R_S) + V_{DS}$$

Solving for  $I_D$ ,

$$I_D = -\frac{V_{DS}}{R_D + R_S} + \frac{E_{DD}}{R_D + R_S} \quad (12.13)$$

Equation (12.13), similar to Eq. (12.5b), is the equation of a load line. Its use is illustrated in the next example.

**example 12.4** Assume that in Fig. 12.14 the sum of  $R_D$  and  $R_S$  is  $40 \Omega$  and  $E_{DD} = 20 \text{ V}$ . If the  $Q$  point is located at  $V_{DSQ} = 10 \text{ V}$ , determine (a) the quiescent current and the gate-source voltage, and (b) the values of  $R_D$  and  $R_S$ . Use the drain characteristics of Fig. 12.15.



**Fig. 12.15** Determining the  $Q$  point of a FET amplifier. (See Example 12.4.)

**solution** (a) Substitution of the given values in Eq. (12.13) yields  $I_D = -V_{DS}/40 + \frac{20}{40} = -V_{DS}/40 + 0.5$ . For  $I_D = 0$ ,  $V_{DS} = 20$  V, which is one point on the load line. For  $V_{DS} = 0$ ,  $I_D = 0.5$  A = 500 mA, the second point on the load line. The load line is drawn between the two points in Fig. 12.15.

For  $V_{DS} = 10$  V, the load line intersects the  $-3$ -volt bias curve. Therefore, the quiescent current  $I_{DQ}$  is approximately 230 mA, and the gate-source bias voltage is  $-3$  V.

(b) Because the gate-source voltage is  $-3$  V,  $-I_D R_S = -3$ . The quiescent drain current was found in (a) to be 230 mA = 0.23 A. Solving for  $R_S$ , we have  $R_S = \frac{3}{0.23} \approx 13 \Omega$ . The value of  $R_D$  is  $40 - 13 = 27 \Omega$ .

**BIASING THE ENHANCEMENT-TYPE MOSFET** An example of biasing an n-channel enhancement-type MOSFET is illustrated in Fig. 12.16. Feedback resistor  $R_F$  is

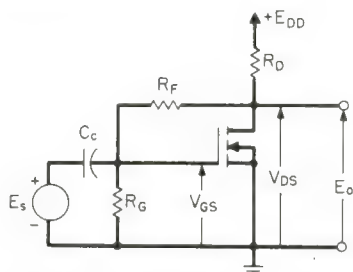


Fig. 12.16 Biasing an n-channel enhancement-type MOSFET.

connected between the drain and gate terminals of the transistor. By voltage division, the gate-source voltage  $V_{GS}$  equals the product of the drain-source voltage  $V_{DS}$  and the ratio  $R_G/(R_G + R_F)$ :

$$V_{GS} = \frac{V_{DS} R_G}{R_G + R_F} \quad (12.14)$$

For a p-channel enhancement-type MOSFET, the polarity of  $E_{DD}$  is reversed.

In addition to providing a bias voltage, the feedback resistor also stabilizes the Q point of the transistor. To minimize the attenuation in signal gain,  $R_F$  may be decoupled as in the common-emitter amplifier using voltage-feedback stabilization.

## 12.5 SMALL-SIGNAL AMPLIFIERS

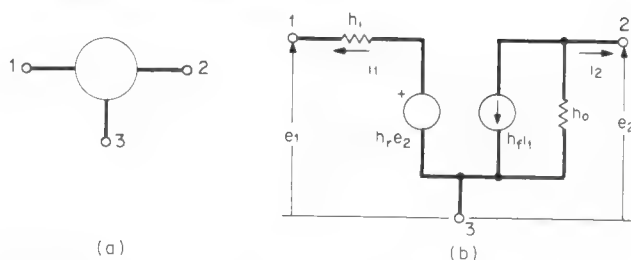
The performance of small-signal amplifiers is determined with the aid of a few simple equations. These equations are based on *models*, or *equivalent circuits*, of the amplifier in question. In this section the performance of small-signal BJT and FET amplifiers is considered. The basic parameters for these devices are defined in Chap. 8.

In the analysis of amplifiers it is desirable to consider specific frequency ranges of signals. This allows the development of simple models and provides insight into amplifier performance. In this section it is assumed that the range of frequencies is less than 5000 Hz, referred to the *midfrequency range*. At midfrequencies, the inherent transistor and stray wiring capacitances may be neglected.

**ANALYZING THE BJT AMPLIFIER** The use of hybrid ( $h$ ) parameters in the analysis of the small-signal BJT amplifier leads to a simple and generalized model. Let the transistor be represented by a circle with three leads numbered 1, 2, and 3, as shown in Fig. 12.17a. For example, lead 1 may be the base, lead 2 the collector, and lead 3 the emitter. A *hybrid model* of the BJT, suitable for midfrequencies, is shown in Fig. 12.17b.

Parameter  $h_i$  is the short-circuit input resistance;  $h_r$  is the reverse transfer ratio;  $h_f$  is the short-circuit forward current gain; and  $h_o$  is the output admittance. The unit for  $h_i$  is ohms and for  $h_o$  it is mhos. Parameters  $h_f$  and  $h_r$  have no units; that is, they are





**Fig. 12.17** Developing a small-signal model of the BJT: (a) General representation of a transistor. (b) Hybrid model.

*numerics.* Sources  $h_r e_2$  and  $h_f i_1$  are examples of *dependent*, or *controlled*, voltage and current sources, respectively (see Chap. 6).

The three useful configurations for the BJT are the common-emitter (CE), common-base (CB), and common-collector (CC), or emitter follower amplifiers. To distinguish the configuration, a second letter is added to each subscript. For example, if the transistor is used in the common-emitter configuration,  $h_i$  becomes  $h_{ie}$ ,  $h_r$  becomes  $h_{re}$ , and so on. The notation is summarized in Table 12.1.

**TABLE 12.1 Hybrid Parameter Notation**

$h$ parameter	Configuration		
	CE	CB	CC
$h_i$	$h_{ie}$	$h_{ib}$	$h_{ic}$
$h_r$	$h_{re}$	$h_{rb}$	$h_{rc}$
$h_f$	$h_{fe}$	$h_{fb}$	$h_{fc}$
$h_o$	$h_{oe}$	$h_{ob}$	$h_{oc}$

Because the CE configuration is widely used, the manufacturer normally provides only common-emitter hybrid parameters on his data sheets. If the common-base or common-collector parameters are needed, they can be obtained by the approximate equations summarized in Table 12.2. In the table the common-base and common-collector parameters are expressed in terms of the common-emitter parameters.

**TABLE 12.2 Approximate Hybrid Parameter Conversions**

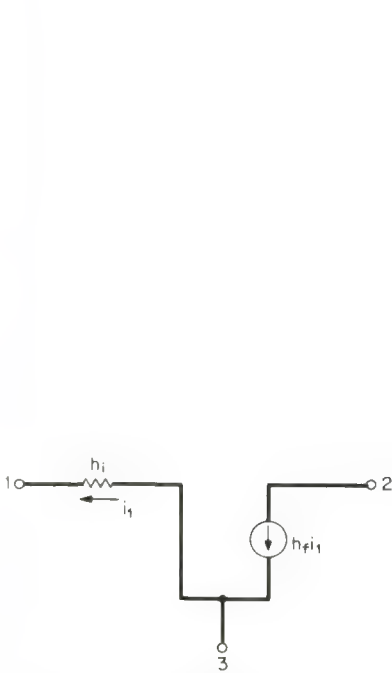
$h_{ib} \approx h_{ie}/(1 + h_{fe})$	$h_{ic} \approx h_{ie}$
$h_{rb} \approx h_{ie} h_{oe}/(1 + h_{fe}) - h_{re}$	$h_{rc} \approx 1$
$h_{fb} \approx -h_{fe}/(1 + h_{fe})$	$h_{fc} \approx -(1 + h_{fe})$
$h_{ob} \approx h_{oe}/(1 + h_{fe})$	$h_{oc} \approx h_{oe}$

**example 12.5** The common-emitter  $h$  parameters of a particular transistor are,  $h_{ie} = 1000 \Omega$ , and  $h_{re} = 10^{-4}$ ,  $h_{fe} = 49$ ,  $h_{oe} = 10^{-5} \text{ mho}$ . Calculate the  $h$  parameters for the transistor in the (a) common-base and (b) common-collector configurations.

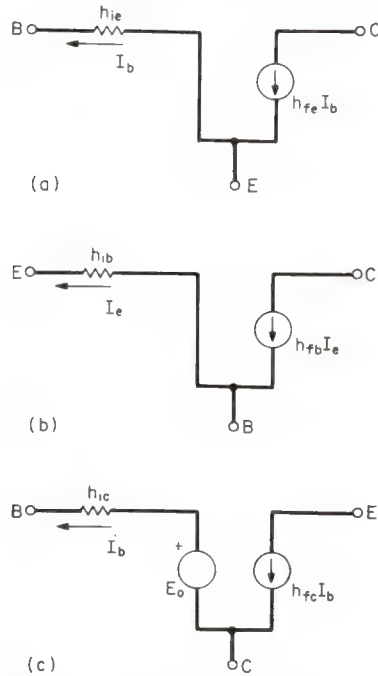
**solution** (a) By the equations of Table 12.2,  $h_{ib} = 1000/(1 + 49) = 20 \Omega$ ;  $h_{rb} = 10^3 \times 10^{-5}/(1 + 49) - 10^{-4} = 2 \times 10^{-4} - 10^{-4} = 10^{-4}$ ;  $h_{fb} = -49/(1 + 49) = -49/50 = -0.98$ ; and  $h_{ob} = 10^{-5}/(1 + 49) = 0.2 \times 10^{-6} \text{ mho}$ .

(b) From Table 12.2,  $h_{ic} = 1000 \Omega$ ;  $h_{rc} = 1$ ;  $h_{fc} = -(1 + 49) = -50$ ; and  $h_{oc} = 10^{-5} \text{ mho}$ .

**Simplified hybrid model** For most small-signal amplifiers, the simplified model of Fig. 12.18 may be used. Comparing Fig. 12.18 with Fig. 12.17b,  $h_r e_2$  and  $h_o$  have been eliminated in the simplified hybrid model. The model for each of the three configurations is illustrated in Fig. 12.19. In each case the numerals have been replaced by letters E, B, and C. It is interesting to note that the same basic model is used for all three configurations.



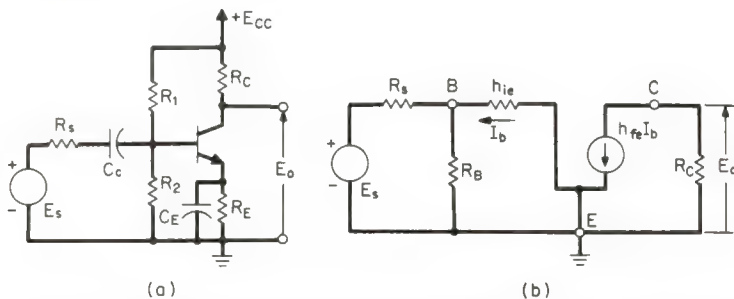
**Fig. 12.18** Simplified hybrid model of a BJT.



**Fig. 12.19** Hybrid models for the three transistor configurations: (a) Common emitter. (b) Common base. (c) Common collector (emitter follower).

**Application of the simplified hybrid model** We consider first the common-emitter amplifier of Fig. 12.20a. Resistance  $R_s$  is the resistance of the source (source resistance). It is reasonable to assume that the reactance of capacitors  $C_c$  and  $C_E$  are zero: that is, they act like short circuits. Because our concern here is with the signal behavior of an amplifier, the dc sources ( $E_{CC}$  in this case) are set to zero. The resultant model of the amplifier is given in Fig. 12.20b. Resistance  $R_B$  is equal to the parallel combination of bias resistors  $R_1$  and  $R_2$ .

Figure 12.21a shows a common-base amplifier, and its model is given in Fig. 12.21b. The emitter follower and its model are illustrated in Fig. 12.22. In deriving the models for these two configuration, the dc sources were set to zero and the coupling capacitor was assumed to act as a short circuit.



**Fig. 12.20** An example of a common-emitter amplifier: (a) Circuit. (b) Model.

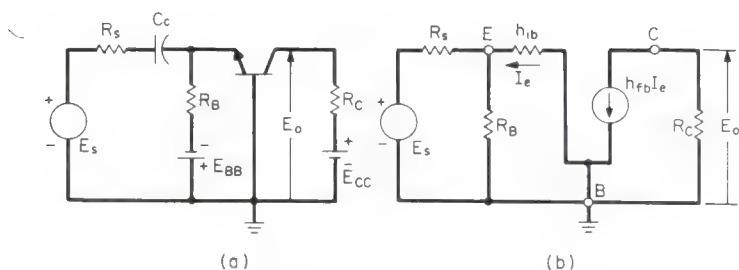


Fig. 12.21 An example of a common-base amplifier: (a) Circuit. (b) Model.

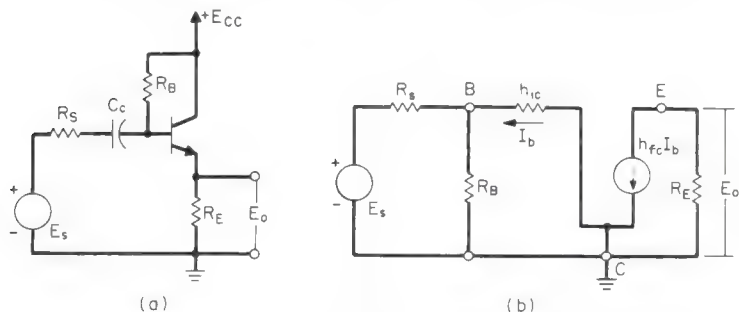


Fig. 12.22 An example of an emitter follower: (a) Circuit. (b) Model.

Equations for current gain  $A_i$ , input resistance  $R_i$ , voltage gain  $A_v$ , and output resistance  $R_o$ , for each of the three configurations are summarized in Table 12.3. The equations, expressed in terms of the common-emitter hybrid parameters, are based on:

1. The simplified hybrid model.
  2. It is assumed that  $R_B$  is much greater than the input resistance and that  $R_C$  is equal to or less than  $5000\ \Omega$ .
- The power gain of an amplifier may be taken as the product of the voltage and current gains.

TABLE 12.3 Expressions for Evaluating the Performance of a BJT Amplifier

	CE	CB	CC
$A_i$	$h_{fe}$	$-h_{fe}/(1 + h_{fe})$	$-(1 + h_{fe})$
$R_i$	$h_{ie}$	$h_{ie}/(1 + h_{fe})$	$h_{ie} + (1 + h_{fe})R_E$
$A_v$	$-h_{fe}R_C/(R_s + R_i)$	$R_C/(R_s + R_i)$	1
$R_o$	$\infty$	$\infty$	$(R_s + h_{ie})/(1 + h_{fe})$

**example 12.6** The transistor used in the amplifiers of Figs. 12.20 to 12.22 have an  $h_{fe}$  of 49 and an  $h_{ie}$  of  $900\ \Omega$ . Assume that  $R_C = 5000$ ,  $R_E = 1000$ , and  $R_s = 100\ \Omega$ . With the aid of Table 12.3, calculate  $A_i$ ,  $R_i$ ,  $A_v$ , and  $R_o$  for each amplifier.

**solution** For the CE amplifier of Fig. 12.20,  $A_i = 49$ ;  $R_i = 900\ \Omega$ ;  $A_v = -49 \times 5000/(100 + 900) = -245$ ; and  $R_o = \infty$ .

For the CB amplifier of Fig. 12.21,  $A_i = -49/50 = -0.98$ ;  $R_i = 900/(1 + 49) = 900/50 = 18\ \Omega$ ;  $A_v = 5000/(100 + 18) = 5000/118 = 42.5$ ; and  $R_o = \infty$ .

For the CC (emitter follower) amplifier of Fig. 12.22,  $A_i = -(1 + 49) = -50$ ;  $R_i = 900 + (1 + 49) \times 1000 = 50\ 900\ \Omega$ ;  $A_v = 1$ ; and  $R_o = (100 + 900)/(1 + 49) = 1000/50 = 20\ \Omega$ .

**Practical considerations** The negative sign before an expression for voltage or current gain indicates that the output signal is  $180^\circ$  out of phase with respect to the input signal. Referring to Fig. 12.23, during the positive half-cycle of the input signal,

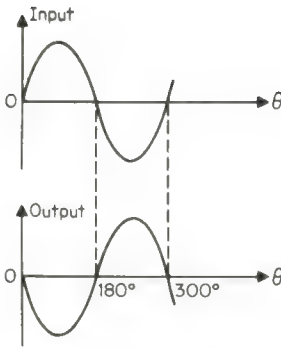


Fig. 12.23 Input and output signal waveforms  $180^\circ$  out of phase.

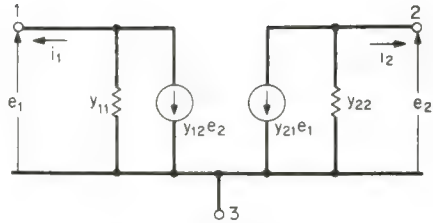


Fig. 12.24 Small-signal admittance ( $y$ ) model of a FET operating at midfrequencies.

the output is a negative half-cycle. As the input signal goes through its negative half-cycle, the output is positive-going.

Although the voltage gain of an emitter follower never exceeds unity, its usefulness lies in its relatively *high input resistance* and *low output resistance*. (In Example 12.6 we found for the emitter follower that  $R_i = 50\,900$  and  $R_o = 20\,\Omega$ .) For example, the emitter follower is used if a voltage source having a large source resistance is required to deliver power to a low-resistance load. If an emitter follower is interposed between the source and load, virtually full power is delivered to the load.

The equations summarized in Table 12.3 assume that  $R_B$  is much greater than the input resistance of the amplifier. If this is not true, because of *current division*, a portion of the input signal current is drained through  $R_B$ , and less base current flows. The method for considering current division is covered in Chap. 6.

If, in Fig. 12.20a, emitter resistor  $R_E$  is not suitably bypassed, the equations for the CE amplifier of Table 12.3 must be modified. For an unbypassed emitter resistor, the input resistance becomes

$$R_i = h_{ie} + (1 + h_{fe})R_E \quad (12.15)$$

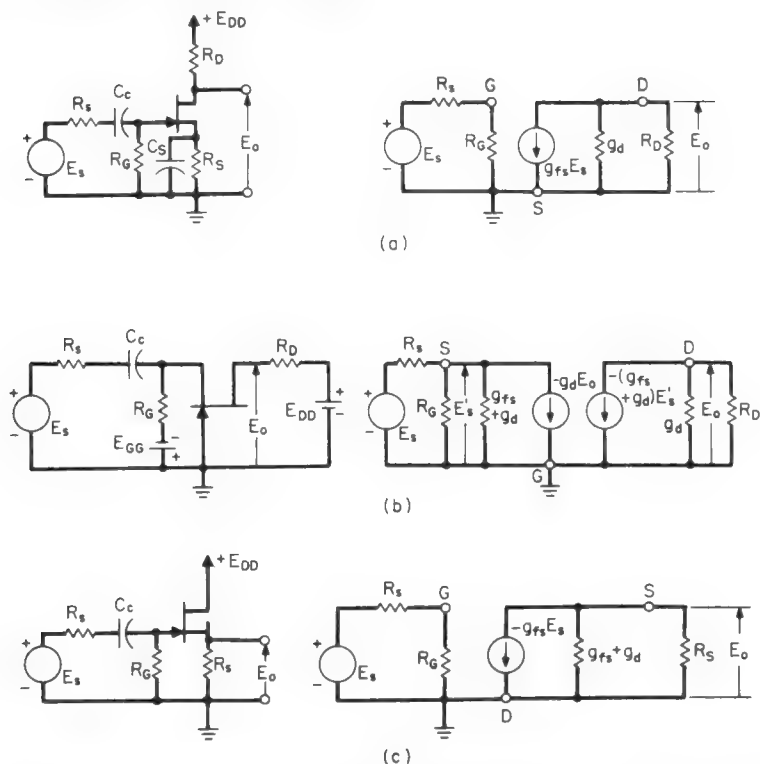
Substitution of Eq. (12.15) for  $R_i$  in the equations for the CE amplifier of Table 12.3 yields the correct results for an unbypassed emitter resistor.

**ANALYZING THE FET AMPLIFIER** In a manner analogous to the development of the small-signal model for the BJT, the basic model for a FET, or MOSFET, operating at midfrequencies is illustrated in Fig. 12.24. For the field-effect device, *admittance* ( $y$ ) *parameters* are used. The equivalent circuit is referred to as a  $y$ , or an admittance, model. Table 12.4 lists the  $y$  parameters for midfrequency operation for the three configurations of the FET or MOSFET: common source (CS), common gate (CG), and common drain (CD), or source follower, amplifiers. Parameter  $g_{fs}$  is the common-source forward transconductance and  $g_d$  the drain-source admittance. Their unit is the *mho*.

The three amplifier configurations and their models for the field-effect transistor are

TABLE 12.4 Small-Signal  $y$  Parameters for the Field-Effect Transistor

Parameter	CS	CG	CD
$y_{11}$	0	$g_{fs} + g_d$	0
$y_{12}$	0	$-g_d$	0
$y_{21}$	$g_{fs}$	$-(g_{fs} + g_d)$	$-g_{fs}$
$y_{22}$	$g_d$	$g_d$	$g_{fs} + g_d$



**Fig. 12.25** Three basic FET amplifiers and their small-signal models: (a) Common source. (b) Common gate. (c) Common drain (source follower).

given in Fig. 12.25. Table 12.5 summarizes the equations for voltage gain, etc., for each circuit. The equations are based on the following assumptions:

1.  $R_G \gg R_s$ .
2.  $g_{fs} \gg g_d$ .
3. Capacitors  $C_c$  and  $C_s$  act like short circuits.
4. Input resistance does not include  $R_G$ .
5. Midfrequency operation.

**example 12.7** A FET has a  $g_{fs} = 0.1$  mho and  $g_d = 10^{-6}$  mho. If  $R_D = 10$  k $\Omega$  and  $R_S = 1$  k $\Omega$ , calculate the values of  $A_v$  and  $R_o$  the common-source, common-gate, and common-drain amplifiers of Fig. 12.25.

**solution** Substituting the given values in the appropriate expressions of Table 12.5, we obtain

**TABLE 12.5** Expressions for Evaluating the Performance of a FET (or MOSFET) Amplifier

	CS	CG	CD
$A_v$	$-g_{fs}/(1/R_L + g_d)$	$g_{fs}/(1/R_L + g_d)$	$g_{fs}/(1/R_L + g_{fs})$
$A_i$	$\infty$	$-1$	$\infty$
$R_i$	$\infty$	$\frac{1}{g_{fs} - g_d g_{fs}(1/R_L + g_d)}$	$\infty$
$R_o$	$1/g_d$	$1/g_d$	$1/g_{fs}$

CS:  $A_v = -g_m/(1/R_L + g_d) = -0.1/(1/10^4 + 10^{-6}) = -0.1/(10^{-4} + 10^{-6}) \approx -0.1/10^{-4} = -1000$ .  $R_o = 1/g_d = 1/10^{-6} = 1 \text{ M}\Omega$ .

CG:  $A_v = g_m/(1/R_L + g_d) = 0.1/(1/10^4 + 10^{-6}) \approx 1000$ .  $R_o = 1/g_d = 1/10^{-6} = 1 \text{ M}\Omega$ .

CD:  $A_v = g_m/(1/R_L + g_m) = 0.1/(1/10^4 + 0.1) \approx 1$ .  $R_o = 1/g_m = 1/0.1 = 10 \Omega$ .

Like the emitter follower, the source follower has a high (infinite) input resistance and a low (10  $\Omega$  in this example) output resistance. Its voltage gain can never exceed unity.

If the emitter resistance of the common-source amplifier (Fig. 12.25a) is not suitably bypassed, the equation for  $A_v$  in Table 12.5 cannot be used. Instead, the voltage gain is expressed by

$$A_v = \frac{-g_m}{1/R_L + g_d + g_m R_s/R_L}$$

## 12.6 FREQUENCY RESPONSE OF AN AMPLIFIER

In the previous section, effects of the inherent transistor capacitances and the coupling capacitor on amplifier performance were neglected. Transistor capacitance attenuates high-frequency signals and the coupling capacitor attenuates low-frequency signals. These effects, for the common-emitter and common-source amplifiers, are examined in this section.

**HIGH-FREQUENCY MODEL OF COMMON-EMITTER AMPLIFIER** The widely used *hybrid-pi model* of a transistor in the CE configuration at high-frequency operation is provided in Fig. 12.26a. Resistance  $r_{bb'}$ , the *base-spreading resistance*, is the ohmic

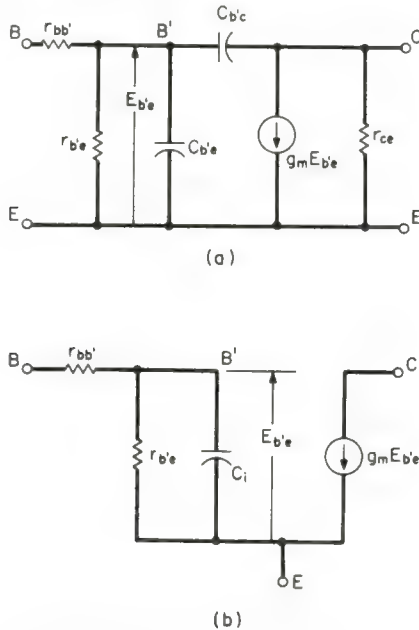


Fig. 12.26 Hybrid-pi model of a BJT in the common-emitter configuration for high-frequency operation: (a) Actual. (b) Simplified.

resistance of the base region. Its value is in the order of 100  $\Omega$ . Base-emitter resistance  $r_{b'e}$  varies with the collector current: the higher the current, the lower its value, and vice versa. The sum of  $r_{bb'}$  and  $r_{b'e}$  is equal to  $h_{ie}$ :

$$h_{ie} = r_{bb'} + r_{b'e} \quad (12.16)$$



## 12-18 Transistor Amplifiers and Oscillators

The dependent source is  $g_m E_{b'e}$ , where  $g_m$  is the *forward transconductance* of the transistor, and  $E_{b'e}$  is the voltage across points  $B'$  and  $E$ . The transconductance is related to  $h_{fe}$  and  $r_{b'e}$  by

$$g_m = \frac{h_{fe}}{r_{b'e}} \quad (12.17a)$$

In general,  $r_{bb'}$  is much less than  $r_{b'e}$ ; hence,  $r_{b'e} \approx h_{ie}$ , and

$$g_m \approx \frac{h_{fe}}{h_{ie}} \quad (12.17b)$$

The transconductance also depends on the dc collector current  $I_C$ . At room temperature (25°C), the relationship is

$$g_m = \frac{I_C}{26} \text{ millimhos} \quad (12.18)$$

where  $I_C$  is expressed in milliamperes.

The *collector-emitter resistance*  $r_{ce}$  is equal to one divided by  $h_{oe}$ :

$$r_{ce} = \frac{1}{h_{oe}} \quad (12.19)$$

It is interesting to note that the noncapacitive elements in the hybrid- $\pi$  model are related to the hybrid parameters.

Capacitances  $C_{b'e}$  and  $C_{b'c}$  are referred to the *diffusion* and *depletion*, or *barrier*, capacitances, respectively. Values of  $C_{b'e}$  range from 100 to 1000 pF (picofarad =  $10^{-12}$  F) and  $C_{b'c}$  is typically a few picofarads.

**Simplified hybrid- $\pi$  model** Owing to a phenomenon, called the *Miller effect*, capacitance  $C_{b'c}$  suitably modified may be placed in parallel with  $C_{b'e}$ . Capacitance  $C_{b'c}$  is multiplied by one plus the product of  $g_m$  and  $R_C$ . The result  $C_M$ , referred to as the *Miller capacitance*, is

$$C_M = (1 + g_m R_C) C_{b'c}$$

Addition of  $C_M$  to  $C_{b'e}$  yields the effective input capacitance  $C_i$ :

$$C_i = C_{b'e} + (1 + g_m R_C) C_{b'c} \quad (12.20)$$

The *simplified hybrid- $\pi$  model* containing  $C_i$  is shown in Fig. 12.26b. Because  $r_{ce}$  is usually much greater than  $R_C$ , it is omitted from the figure. The high-frequency model of the common-emitter amplifier of Fig. 12.20a appears as in Fig. 12.27.

Because at increasing frequencies the reactance of  $C_i$  decreases, the voltage across it,  $E_{b'e}$ , also drops. As a result, the collector signal current  $I_c$ , which equals the product of  $g_m$  and  $E_{b'e}$ , and the output voltage  $E_o$  are reduced. The voltage gain, therefore, falls with increasing frequency.

The upper -3-dB frequency  $f_H$  at which the voltage gain falls to approximately 70 percent of its value at midfrequencies (usually taken at 1 kHz) is determined as follows. Signal source  $E_s$  is set to zero. (As explained in Chap. 6, in setting a voltage source to zero, it is short-circuited; in setting a current source to zero, it is open-circuited.) A

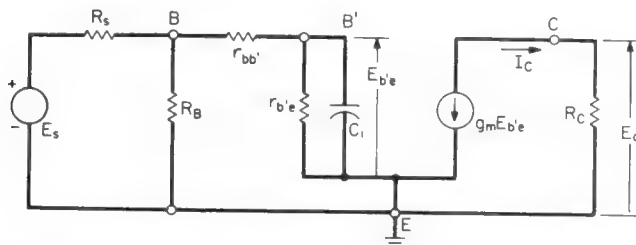


Fig. 12.27 High-frequency model of a common-emitter amplifier.

reasonable assumption is that  $R_B$  is much greater than source resistance  $R_s$ . Therefore, neglecting  $R_B$ , the equivalent resistance  $R_{eq}$  across  $C_i$  is  $(R_s + r_{bb'})$  in parallel with  $r_{b'e}$  (Fig. 12.28). Hence,

$$R_{eq} = \frac{(R_s + r_{bb'})r_{b'e}}{R_s + r_{bb'} + r_{b'e}} \quad (12.21)$$

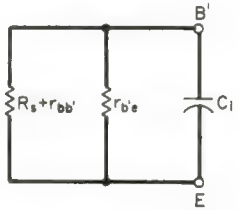
At the upper  $-3$ -dB frequency  $f_H$ ,  $R_{eq}$  is equal to the reactance of  $C_i = 1/2\pi f_H C_i$ . Therefore,

$$R_{eq} = \frac{1}{2\pi f_H C_i}$$

Solving for  $f_H$ ,

$$f_H = \frac{1}{2\pi R_{eq} C_i} = \frac{0.159}{R_{eq} C_i} \quad (12.22)$$

since  $1/2\pi = 0.159$ .



**Fig. 12.28** Model of input circuit for determining the upper  $-3$ -dB frequency,  $f_H$ .

**example 12.8** In Fig. 12.27,  $g_m = 0.1$  mho,  $C_{b'e} = 300$  pF,  $C_{b'c} = 4$  pF,  $R_C = 1000 \Omega$ ,  $R_s = 900 \Omega$ ,  $r_{bb'} = 100 \Omega$ , and  $r_{b'e} = 1000 \Omega$ . Neglecting  $R_B$ , calculate (a)  $C_i$ , (b)  $R_{eq}$ , and (c)  $f_H$ .

**solution** (a) Substituting the given values, by Eq. (12.20),  $C_i = 300 + (1 + 0.1 \times 1000) \times 4 = 300 + 101 \times 4 = 704$  pF.

(b) By Eq. (12.21),  $R_{eq} = (900 + 100) \times 1000 / (900 + 100 + 1000) = 500 \Omega$ .

(c) By Eq. (12.22),  $f_H = 0.159 / (500 \times 704 \times 10^{-12}) = 452,000$  Hz = 452 kHz = 0.452 MHz.

**Figures of merit** Three commonly used figures of merit for comparing the high-frequency performance of bipolar transistors are:

1. **Short-circuit beta cutoff frequency**  $f_\beta$  is the frequency at which  $h_{fe}$  is equal to approximately 70 percent of its value at midfrequencies. It is expressed by

$$f_\beta = \frac{0.159}{r_{b'e}(C_{b'e} + C_{b'c})} \quad (12.23)$$

The lower the values of  $r_{b'e}$ ,  $C_{b'e}$ , and  $C_{b'c}$ , the higher is  $f_\beta$ .

2. **Short-circuit gain-bandwidth product**  $f_T$  is the frequency at which  $h_{fe}$  is equal to one. It is given by

$$f_T = \frac{0.159 g_m}{C_{b'e} + C_{b'c}} \quad (12.24)$$

The relationship of  $h_{fe}$ ,  $f_T$ , and  $f_\beta$  is given by

$$f_T = h_{fe} f_\beta \quad (12.25)$$

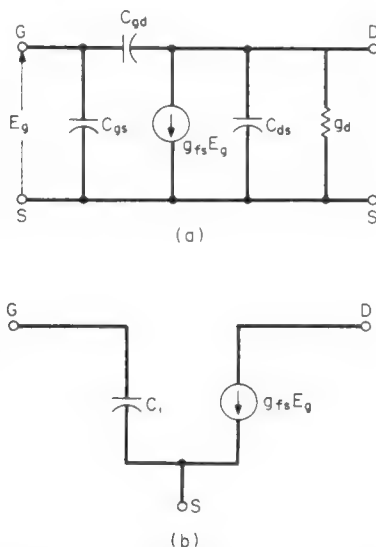
3. **Short-circuit alpha cutoff frequency**  $f_\alpha$  is the frequency at which  $h_{fb}$  is equal to approximately 70 percent of its value at midfrequencies. It is expressed by

$$f_\alpha = \frac{0.159 g_m}{C_{b'e}} \quad (12.26)$$

Because  $C_{b'c}$  is generally much less than  $C_{b'e}$ , comparing Eqs. (12.24) and (12.26),  $f_\alpha \approx f_T$ .

**HIGH-FREQUENCY MODEL OF COMMON-SOURCE AMPLIFIER** A high-frequency model of a FET in the common-source configuration is given in Fig. 12.29a. Capacitances  $C_{gs}$  and  $C_{gd}$  are depletion capacitances between gate and source, and gate and drain, respectively. Capacitance  $C_{ds}$  is the drain-source capacitance of the channel.

By using the Miller effect, as was done for the common-emitter amplifier, the high-



**Fig. 12.29** High-frequency model of a FET (or MOSFET) in the common-source configuration: (a) Actual. (b) Simplified.

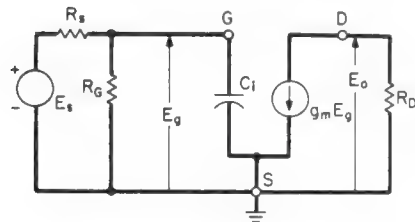
frequency model is simplified, as shown in Fig. 12.29b. Input capacitance  $C_i$  is expressed by

$$C_i = C_{gs} + (1 + g_m R_D) \quad (12.27)$$

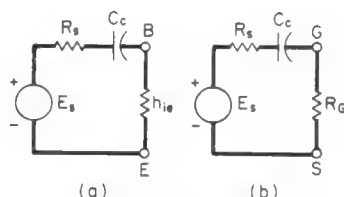
The value of  $1/g_d$  is generally much greater than practical values of load resistance  $R_D$ ; therefore it is omitted from the simplified model. Because  $C_i$  is the influencing capacitance in determining the high-frequency response,  $C_{ds}$  is also omitted from the figure.

The high-frequency model of the common-source amplifier of Fig. 12.25a is illustrated in Fig. 12.30. Similar to the analysis of the CE amplifier, the upper -3-dB frequency  $f_H$  is determined by equating source resistance  $R_s$  to the reactance of  $C_i$ :

$$R_s = \frac{0.159}{f_H C_i}$$



**Fig. 12.30** High-frequency model of a common-source amplifier.



**Fig. 12.31** Input circuits at low frequency for (a) a CE amplifier and (b) a CS amplifier.

Solving for  $f_H$ ,

$$f_H = \frac{0.159}{R_s C_i} \quad (12.28)$$

**LOW-FREQUENCY RESPONSE** Assume that in a common-emitter amplifier the emitter resistor is suitably bypassed. (For a common-source amplifier, it is assumed that the source resistor is suitably bypassed.) The low-frequency response is then determined by coupling capacitor  $C_c$ . This is illustrated in Fig. 12.31 where the input circuits for the CE and CS amplifiers at low-frequency operation are shown.

The reactance of a capacitor, as mentioned earlier, decreases with increasing frequencies and increases with decreasing frequencies. Thus, at high frequencies, the reactance is so much less than the series resistance in the circuit that it acts like a short circuit. As the frequency is reduced, however, the reactance is increased. As a result, the voltage across the base and emitter (or gate and source) terminals is reduced and the gain of the amplifier falls.

The lower -3-dB frequency  $f_L$  is determined by setting the total series resistance in the input circuit equal to the reactance of the coupling capacitor at  $f_L$ . For the CE amplifier,

$$R_s + h_{ie} = \frac{0.159}{f_L C_c}$$

Solving for  $f_L$ ,

$$f_L = \frac{0.159}{(R_s + h_{ie})C_c} \quad (12.29a)$$

For the CS amplifier,

$$R_s + R_G = \frac{0.159}{f_L C_c}$$

and

$$f_L = \frac{0.159}{(R_s + R_G)C_c} \quad (12.29b)$$

A similar approach is used in calculating the low-frequency response of the other amplifier configurations.

## 12.7 CASCADING AMPLIFIER STAGES

To realize greater signal gain than obtainable from a single amplifier stage, a number of stages are *cascaded*, as illustrated in the block diagram of Fig. 12.32. In a cascaded amplifier, the output of one amplifier stage is *coupled* to the input of the following stage. If the stages are *noninteracting*, that is, if there is no appreciable change in the gain or other parameters of an individual stage when cascaded, the overall gain  $A_t$  is equal to the product of the individual gains,

$$A_t = A_1 \times A_2 \times \cdots \times A_n \quad (12.30)$$

where  $A_1$  is the gain of the first stage,  $A_2$  the gain of the second stage, and  $A_n$  the gain of the  $n$ th stage.

Owing to its nearly infinite input resistance, the FET amplifier may be considered as being noninteracting. The BJT, because its input resistance is relatively low, interacts and Eq. (12.30) cannot be used. Instead, working from the output toward the input stage, the gain of each stage must be considered separately, taking into account loading effects of the cascaded stages.

**COUPLING CASCADED STAGES** Two basic methods are used for coupling cascaded stages:

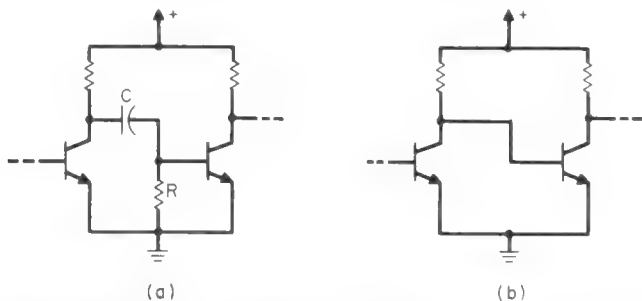
1. Resistance-capacitance (RC) coupling.
2. Direct coupling (DC).



Fig. 12.32 Block diagram of a cascaded amplifier.

Examples of resistance-capacitance and direct-coupled BJT stages are illustrated in Fig. 12.33. (If a FET is substituted for a BJT, the circuits are identical.)

In *RC coupling*, the dc collector voltage of the preceding stage is blocked by the coupling capacitor from the base of the following stage. In *direct coupling*, the collector of one transistor is coupled directly to the base of the following transistor. Because the dc operating point of a transistor tends to drift, special circuit precautions are taken in *DC amplifier design*. The greatest use of direct coupling is in integrated-circuit amplifiers (see Chap. 13). Unless direct current or slowly varying signals are to be amplified, *RC coupling* is invariably used in discrete amplifiers.

Fig. 12.33 Coupling cascaded stages: (a) Resistance-capacitance (*RC*) coupling. (b) Direct coupling (*DC*).

**FREQUENCY RESPONSE OF A CASCADED AMPLIFIER** The overall frequency response of a cascaded amplifier is less than that for an individual stage. For *noninteracting identical stages*, the upper  $-3$ -dB frequency of an  $n$ -stage cascaded amplifier  $f'_H$  is

$$f'_H = f_H \sqrt{2^{1/n} - 1} \quad (12.31a)$$

The lower  $-3$ -dB frequency  $f'_L$  is

$$f'_L = f_L / \sqrt{2^{1/n} - 1} \quad (12.31b)$$

where  $f_H$  and  $f_L$  are the upper and lower  $-3$ -dB frequencies of an individual stage.

**example 12.9** An individual stage of a two-stage *RC*-coupled amplifier has  $f_H = 20$  kHz and  $f_L = 200$  Hz. Its bandwidth, therefore, is  $20\,000 - 200 = 19\,800$  Hz = 19.8 kHz. Assuming that the stages are identical and noninteracting, calculate the bandwidth of the cascaded amplifier.

**solution** By Eq. (12.31a),  $f'_H = 20 \sqrt{2^{1/2} - 1}$ . The quantity  $2^{1/2} = \sqrt{2} = 1.414$ ; hence,  $\sqrt{1.414 - 1} = \sqrt{0.414} = 0.64$ . Therefore,  $f'_H = 20 \times 0.64 = 12.8$  kHz. By Eq. (12.31b),  $f'_L = 200/0.64 = 313$  Hz.

The bandwidth of the cascaded amplifier is  $f'_H - f'_L = 12\,800 - 313 \approx 12.5$  kHz. With respect to an individual stage, the bandwidth of the cascaded amplifier was reduced in excess of 60 percent.

## 12.8 POWER AMPLIFIERS

A useful criterion for comparing power amplifiers is the conversion efficiency  $\eta$  (Greek letter *eta*). The conversion efficiency is defined as the signal power delivered to the load  $P_{ac}$  to the dc power  $P_{dc}$  supplied to the collector or drain circuit of a transistor. Expressed as a percentage,

$$\eta = \frac{P_{ac}}{P_{dc}} \times 100\% \quad (12.32)$$

For example, the maximum conversion efficiency realized with class A operation is 50 percent and with class B operation 78.4 percent. Class A and B power amplifiers are examined in the following sections.

**CLASS A POWER AMPLIFIER** A typical class A power amplifier is illustrated in Fig. 12.34. Load  $R_L$ , which may be the coil of a loudspeaker, is coupled to the collector circuit of the transistor by *coupling transformer*  $T$ . The turns ratio of the transformer  $n$  is the ratio of  $N_1$  to  $N_2$  turns:

$$n = \frac{N_1}{N_2} \quad (12.33)$$

The value of  $R_L$  "seen" by the transistor is modified. The modified, or *reflected*, value  $R'_L$  is given by the product of  $n$  squared and  $R_L$ :

$$R'_L = n^2 R_L \quad (12.34)$$

The output power is generally determined graphically. Two load lines are drawn on the collector (or drain) family of curves. The dc, or *static*, load line represents the dc resistance of transformer winding  $N_1$ . Because its resistance is extremely low, it is taken to be zero. A resistance value of zero ohms appears as a *vertical line* perpendicular to the voltage axis at  $E_{CC}$  (or  $E_{DD}$ ), as shown in Fig. 12.35.

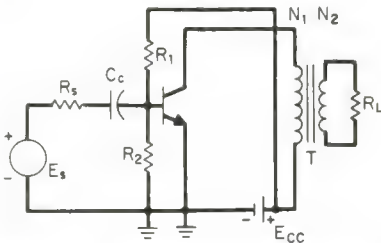


Fig. 12.34 Typical class A power amplifier.

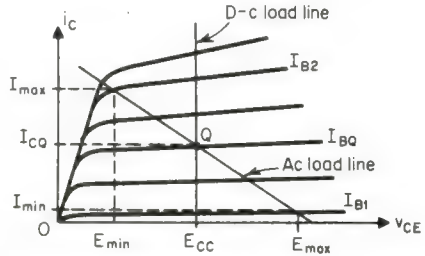


Fig. 12.35 Superimposing the dc and ac load lines on the collector characteristics.

The  $Q$  point is determined by the intersection of the dc load line and base current  $I_{BQ}$ . The value of the base current depends on the values of  $R_1$  and  $R_2$  of the biasing network (Fig. 12.34). Intersecting the  $Q$  point and having a slope of  $-1/R'_L$  is the ac, or *dynamic*, load line.

Assume that the input sine wave signal has a positive peak value of  $I_{B2}$  and a negative peak value of  $I_{B1}$ . The corresponding maximum and minimum values of collector current and voltage are found by the intersection of the ac load line and the  $I_{B1}$  and  $I_{B2}$  curves. As indicated in Fig. 12.35, these values are denoted by  $I_{max}$ ,  $E_{max}$ ,  $I_{min}$ , and  $E_{min}$ , respectively. The output power  $P_{ac}$  in watts is

$$P_{ac} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8} \quad (12.35)$$

**COLLECTOR DISSIPATION** The power dissipated as heat in the collector is referred to as collector dissipation  $P_C$ . (For a FET, the dissipated power in the drain is called the *drain dissipation*  $P_D$ .) The value of  $P_C$  is

$$P_C = E_{CC} I_{CQ} - P_{ac} \quad (12.36a)$$

A class A amplifier operating without a signal is said to be *idling*. When idling,  $P_{ac} = 0$ , and the amplifier is *dissipating maximum power*,  $P_{C,max}$ :

$$P_{C,max} = E_{CC} I_{CQ} \quad (12.36b)$$

It is important that the maximum dissipated power never exceed the value given by the manufacturer. Further, from Fig. 12.35, it is seen that the maximum collector signal



## 12-24 Transistor Amplifiers and Oscillators

voltage  $E_{\max}$  exceeds  $E_{CC}$ . The value of  $E_{\max}$  must always be less than the collector (or drain) breakdown voltage of the transistor specified by the manufacturer.

**SECOND-HARMONIC DISTORTION** Second-harmonic distortion of an amplifier provides an indication of how faithful the amplified output signal is with respect to the input signal. Using the values found graphically in determining the output power, second-harmonic distortion  $D_2$  is

$$D_2 = \frac{|I_{\max} + I_{\min} - 2I_{CQ}|}{2|I_{\max} - I_{\min}|} \times 100\% \quad (12.37)$$

The vertical lines indicate the *magnitude of the quantity*.

**example 12.10** The dc and ac load lines are superimposed on the collector characteristics of Fig. 12.36. The Q point for class A operation is located at approximately  $E_{CC} = 4$  V and  $I_{CQ} = 1$  A. For an input signal swing between  $I_B = 200$  mA and 10 mA, the following values are estimated from the intersection of the ac load line and the base current curves.  $E_{\max} \approx 7.3$  V,  $E_{\min} \approx 2.7$  V,  $I_{\max} \approx 1.4$  A, and  $I_{\min} \approx 0.2$  A. Find (a)  $P_{ac}$ ; (b)  $\eta$ ; (c)  $P_C$  and  $P_{C,\max}$ ; (d)  $D_2$ .

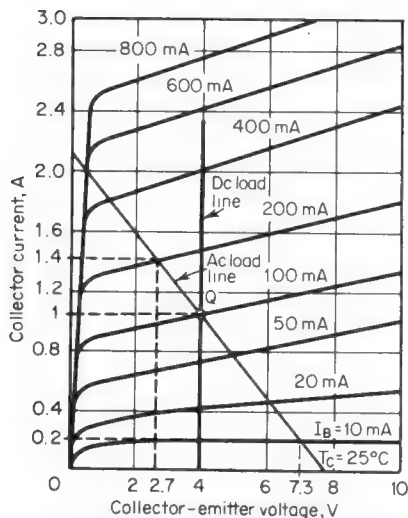
**solution** (a) By Eq. (12.35),  $P_{ac} = (7.3 - 2.7)(1.4 - 0.2)/8 = 4.6 \times 1.2/8 \approx 0.69$  W.

(b)  $P_{dc} = E_{CC}I_{CQ} = 4 \times 1 = 4$  W. By Eq. (12.32),  $\eta = (0.69/4) \times 100\% = 17.2\%$ . This is considerably less than the maximum conversion efficiency of 50 percent for class A operation.

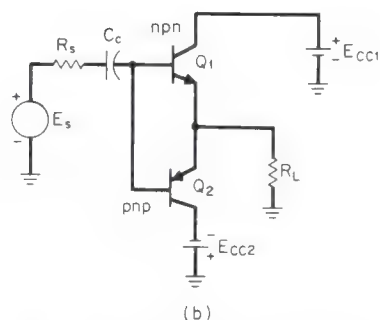
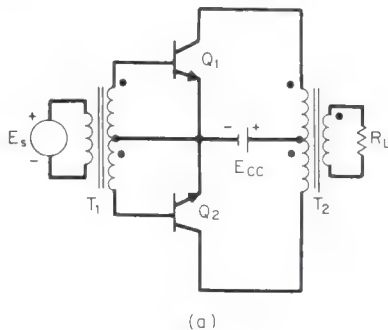
(c) By Eqs. (12.36a) and (12.36b),  $P_C = 4 - 0.69 = 3.31$  W;  $P_{C,\max} = 4$  W.

(d) By Eq. (12.37),  $D_2 = \frac{|1.4 + 0.2 - 2 \times 1|}{2|1.4 - 0.2|} \times 100\% = 0.4/2.4 \times 100\% = 16.6\%$ .

**PUSH-PULL POWER AMPLIFIERS** Examples of push-pull amplifier circuits are illustrated in Fig. 12.37. An advantage of push-pull over single-ended operation is that,



**Fig. 12.36** Determining output power of a class A single-ended amplifier. (See Example 12.10.)



**Fig. 12.37** Examples of class B push-pull power amplifiers: (a) Conventional type. (b) Complementary symmetry type.

assuming identical transistors, all even harmonics of distortion are canceled. If the transistors are operated as class B, conversion efficiencies as high as 78.4 percent are realized. Further, when the amplifier is idling, the collector dissipation is zero. The discussion to follow is limited to class B operation.

The circuit of Fig. 12.37a requires two center-tapped transformers,  $T_1$  and  $T_2$ . Transformer  $T_1$  is referred to as a *phase-inverting input transformer*. The voltage polarity at the dotted terminals is always the same, either plus or minus. If the base of transistor  $Q_1$  is positive, the base of transistor  $Q_2$  is negative, and vice versa. Output transformer  $T_2$  couples load resistance  $R_L$  across the collectors of  $Q_1$  and  $Q_2$ .

Assume that the signal is a sine wave. On the positive half-cycle, the base of  $Q_1$  is positive and the base of  $Q_2$  negative. Transistor  $Q_1$  conducts and  $Q_2$  is cut off. During the negative half-cycle, the polarities are reversed. Now,  $Q_1$  is cut off and  $Q_2$  conducts current. During a cycle, therefore, each transistor in a class B push-pull amplifier conducts current half the time.

The circuit of Fig. 12.37b, referred to as a *complementary symmetry amplifier*, does not require input and output transformers. (It may be regarded as a *transformerless* power amplifier.) Transistor  $Q_1$  is npn and  $Q_2$  is pnp. For good operation,  $Q_1$  and  $Q_2$  should be closely matched in their electrical characteristics. During the positive half-cycle,  $Q_1$  conducts and  $Q_2$  is cut off. On the negative half-cycle the situation is reversed:  $Q_2$  conducts and  $Q_1$  is cut off.

**CROSSOVER DISTORTION** The input resistance of an amplifier increases with decreasing collector current and decreases with increasing collector current. Consequently, in class B operation, the sine wave is flattened, as illustrated in Fig. 12.38. The flattening is referred to as *crossover distortion*.

To reduce crossover distortion, power amplifiers are generally operated as class AB. Under this mode of operation, a small collector current flows for zero signal current. The conversion efficiency of a class AB push-pull amplifier is in the order of 60 to 70 percent.

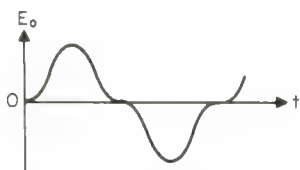


Fig. 12.38 Illustration of crossover distortion.

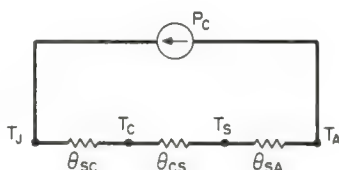


Fig. 12.39 Thermal equivalent circuit of a transistor mounted on a heat sink.

**HEAT SINKS** Because of possible large collector dissipation, power transistors (as well as power diodes and other semiconductor devices) are often mounted on heat sinks. A heat sink is a metal plate or structure, typically aluminum, which aids in dissipating the heat away from the device.

One can draw a *thermal equivalent circuit* of a transistor mounted on a heat sink, as shown in Fig. 12.39. This circuit is similar to a series circuit of resistors where temperature  $T$  is analogous to voltage; thermal resistance  $\Theta$  (Greek letter *theta*) is analogous to electrical resistance; and dissipated power  $P_C$  is analogous to current. By Ohm's law,  $I = E/R$ ; for the thermal equivalent circuit, one can write  $P_C = T/\Theta$ .

The temperature of the collector-base junction is designated by  $T_J$ ;  $T_C$  is the temperature of the transistor case;  $T_S$  is the temperature of the heat sink; and  $T_A$  is the ambient (environment) temperature. The unit of temperature is degrees Celsius.

Thermal resistance  $\Theta_{JC}$  is the resistance to the flow of heat between the junction and case of the device;  $\Theta_{CS}$  is the thermal resistance between the case and heat sink; and  $\Theta_{SA}$  is the thermal resistance between the heat sink and the ambient. The unit for  $\Theta$  is degrees Celsius per watt and  $P_C$  is in watts.

Applying Ohm's law to the thermal equivalent circuit, the allowable dissipated power

## 12-26 Transistor Amplifiers and Oscillators

is equal to the difference in the junction and ambient temperature divided by the sum of the thermal resistances:

$$P_c = \frac{T_J - T_A}{\Theta_{JC} + \Theta_{CS} + \Theta_{SA}} \quad (12.38)$$

The application of Eq. (12.38) is illustrated in the following example.

**example 12.11** A silicon transistor is mounted on a heat sink. If  $T_J = 125^\circ\text{C}$  and  $T_A = 25^\circ\text{C}$ , determine the thermal resistance of the heat sink needed in order for the transistor to dissipate 50 W. Assume that  $\Theta_{JC} = 0.5^\circ\text{C/W}$  and  $\Theta_{CS} = 0.3^\circ\text{C/W}$ .

**solution** Substitution of the given values in Eq. (12.38) yields  $50 = (125 - 25)/(0.5 + 0.3 + \Theta_{SA}) = 100/(0.8 + \Theta_{SA})$ . Solving,  $0.8 + \Theta_{SA} = 100/50 = 2$ . Hence,  $\Theta_{SA} = 2 - 0.8 = 1.2^\circ\text{C/W}$ .

## 12.9 FEEDBACK AMPLIFIERS

Examples of feedback have already been considered in this chapter. It was seen that current- and voltage-feedback stabilization helped maintain the  $Q$  point of an amplifier relatively fixed, regardless of changes in temperature or other parameters. In general, an amplifier with feedback (*feedback amplifier*) provides gain stability, increases frequency response, and decreases distortion.

The operation of a feedback amplifier is influenced by its output signal, in addition to its input signal. This is realized by returning, or feeding back, a portion or all of the output signal to the input of the amplifier. In *voltage feedback*, the feedback signal is proportional to the load voltage; in *current feedback*, the returned signal is proportional to the load current.

If the feedback signal subtracts from the input signal, *negative feedback* is said to exist. For *positive feedback*, the feedback signal adds to the input signal. Negative feedback is used in amplifiers; positive feedback is the basis for the operation of sinusoidal oscillators.

**PROPERTIES OF NEGATIVE FEEDBACK** To highlight some important properties of negative feedback, reference is made to the block diagram of a feedback amplifier in Fig. 12.40. Output voltage  $E_o$ , which is across load resistor  $R_L$ , is also impressed across the input terminals of the feedback network. The output voltage of the feedback network  $E_f$  is connected in series with signal source  $E_s$ . The gain of the amplifier block is  $A_v$ , and the gain of the feedback network is  $B$ .

Because the feedback voltage is proportional to the output voltage, voltage feedback is obtained. Furthermore, the feedback voltage is in series with the input signal. For these reasons, the configuration of Fig. 12.40 is referred to as a *voltage-feedback series-input feedback amplifier*.

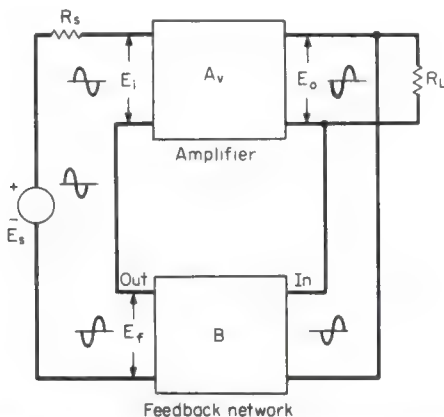


Fig. 12.40 Block diagram of a feedback amplifier.

For the verification that negative feedback indeed exists, voltage waveforms at the input and output terminals of the amplifier and feedback network are indicated in Fig. 12.40. Assume that the input sine wave to the amplifier  $E_i$  is positive-going. Also, the numerical value of  $A_v$  is negative, indicating that the output is  $180^\circ$  out of phase with respect to the input. As a result, output  $E_o$  is negative-going.

The input to the feedback network, which is equal to  $E_o$ , is also negative-going. Because the feedback network is generally comprised of resistors, there is no phase inversion, and its output  $E_f$  is negative-going. Negative feedback exists since  $E_f$  opposes  $E_i$ .

We now derive an expression for voltage gain with feedback  $A_{vf}$ . Referring back to Fig. 12.40, input voltage  $E_i$  is equal to the algebraic sum of the signal voltage  $E_s$  and feedback voltage  $E_f$ :

$$E_i = E_s + E_f \quad (12.39a)$$

The feedback voltage is equal to the product of output voltage  $E_o$  and the gain of the feedback network  $B$ :

$$E_f = BE_o \quad (12.39b)$$

Substitution of Eq. (12.39b) in Eq. (12.39a) yields

$$E_i = E_s + BE_o \quad (12.39c)$$

The gain of the amplifier block  $A_v$  is equal to the ratio of its output to input voltages

$$A_v = \frac{E_o}{E_i} \quad (12.39d)$$

Substituting Eq. (12.39c) for  $E_i$  in Eq. (12.39d) results in

$$A_v = \frac{E_o}{E_s + BE_o} \quad (12.39e)$$

Solving Eq. (12.39e) for  $A_{vf}$ , which is equal to output voltage  $E_o$  divided by the signal voltage  $E_s$ , we obtain

$$A_{vf} = \frac{A_v}{1 - BA_v} \quad (12.40)$$

It is common practice to refer to  $A_{vf}$  as the *closed-loop gain*,  $A_v$  as the *open-loop gain*, and the product  $BA_v$  as the *loop gain*.

Equation (12.40) illustrates some very important facts about feedback. If  $B = 0$  (no feedback present), the closed-loop and open-loop gains are equal ( $A_{vf} = A_v$ ). For negative feedback, the product  $BA_v$  is negative. Consequently, the denominator is greater than one and the voltage gain with negative feedback is less than without feedback. This is a very important property of negative feedback.

If  $BA_v$  is much greater than one, the one can be neglected in the denominator of Eq. (12.40). Consequently,

$$A_{vf} \approx \frac{A_v}{-BA_v}$$

The  $A_v$  terms cancel and

$$A_{vf} \approx -\frac{1}{B} \quad (12.41)$$

Equation (12.41) demonstrates that if  $BA_v \gg 1$ , the voltage gain depends essentially on the gain of the feedback network. This means that, as long as  $B$  is constant, even if the gain of the amplifier changes owing to variation in temperature or in device parameters, the gain of the feedback amplifier is hardly affected. The gain of the feedback network, because it is usually comprised of resistors, is easy to stabilize. Another important property of negative feedback, therefore, is *gain stability*.

**example 12.12** Negative feedback is to be introduced around an amplifier with an open-loop voltage gain of  $A_v = -200$ . It is required that the voltage gain with feedback also equal

## 12-28 Transistor Amplifiers and Oscillators

-200 ( $A_{vf} = -200$ ). Because of negative feedback, the open-loop gain must be increased to realize  $A_{vf} = -200$ . If  $B = 4 \times 10^{-3}$ , determine the new value of  $A_r$ .

**solution** Solving Eq. (12.40) for  $A_r$  yields

$$A_r = \frac{A_{vf}}{1 + BA_{vf}} \quad (12.42)$$

Substituting the given values in Eq. (12.42), we have  $A_r = -200/(1 - 4 \times 10^{-3} \times 200) = -200/0.2 = -1000$ .

**Bandwidth** Another advantage of negative feedback is that it extends the bandwidth of an amplifier. Letting the upper -3-dB frequency with feedback be  $f_{HF}$  and the lower -3-dB frequency be  $f_{LF}$ , we have

$$f_{HF} = f_H(1 - BA_r) \quad (12.43a)$$

$$f_{LF} = \frac{f_L}{1 - BA_r} \quad (12.43b)$$

For example, if  $f_H = 10$  kHz and  $f_L = 100$  Hz, then for  $1 - BA_r = 10$ ,  $f_{HF} = 10 \times 10$  kHz = 100 kHz and  $f_{LF} = 100/10 = 10$  Hz. The bandwidth has been extended by a factor slightly greater than ten.

**Distortion** Negative feedback also reduces distortion in an amplifier. Letting  $D_F$  equal the distortion with feedback, and  $D$  the distortion without feedback, then

$$D_F = \frac{D}{1 - BA_r} \quad (12.44)$$

For example, if the second-harmonic distortion without feedback  $D_2$  is 10 percent, then the second-harmonic distortion with feedback  $D_{F2} = 10/10$  is 1 percent. The distortion has been reduced by a factor of ten.

**Instability** At very low and high frequencies, the loop gain  $BA_r$  may become equal to, or close to, one. If a one is substituted for  $BA_r$  in Eq. (12.40), we obtain

$$A_{vf} = \frac{A_r}{1 - 1} = \frac{A_r}{0} = \infty$$

The gain of the feedback amplifier, therefore, approaches infinity. What this means in practical terms is that the amplifier is *unstable* and sine-wave oscillations are generated. This is undesirable for amplifiers although, as we shall see, a necessary condition for a sine-wave oscillator.

Whenever negative feedback is introduced in an amplifier, an opportunity exists for the feedback amplifier to become unstable. There are criteria, such as the *Nyquist criterion*, for determining the stability of a feedback amplifier. *Compensation networks* can be used to convert an unstable feedback amplifier to a stable one.

**EFFECTS ON INPUT AND OUTPUT RESISTANCE** In addition to providing voltage and current gain stability, increased bandwidth, and reduced distortion, negative feedback also modifies the input and output resistances of an amplifier. Depending on how feedback is introduced, it can either decrease or increase the input and output resistances of a feedback amplifier. There are four basic types of feedback, as illustrated in the examples of Fig. 12.41.

a. *Voltage feedback, series input.* Input resistance,  $R_{iF}$ , is *greater* than input resistance without feedback,  $R_i$ :  $R_{iF} > R_i$ . Output resistance,  $R_{oF}$ , is *less* than output resistance without feedback,  $R_o$ :  $R_{oF} < R_o$ .

b. *Voltage feedback, shunt input:*  $R_{iF} < R_i$ ;  $R_{oF} < R_o$ .

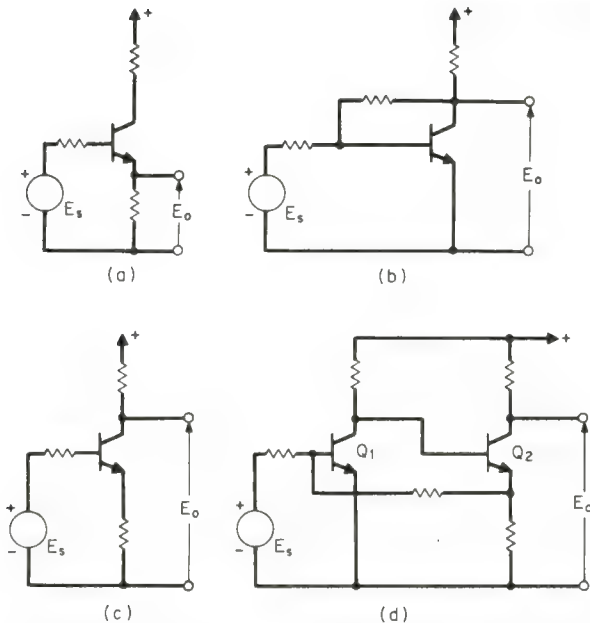
c. *Current feedback, series input:*  $R_{iF} > R_i$ ;  $R_{oF} > R_o$ .

d. *Current feedback, shunt input:*  $R_{iF} < R_i$ ;  $R_{oF} > R_o$ .

## 12.10 SINUSOIDAL OSCILLATORS

In the discussion of instability in feedback amplifiers, it was pointed out that if the loop gain  $BA_r = 1$ , the amplifier oscillates. The condition that the loop gain be unity for





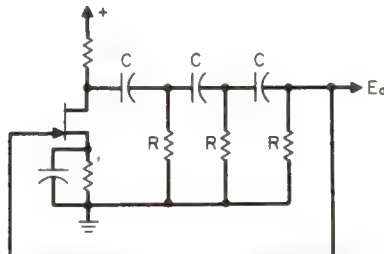
**Fig. 12.41** Four basic types of feedback: (a) Voltage feedback, series input. (b) Voltage feedback, shunt input. (c) Current feedback, series input. (d) Current feedback, shunt input.

sine-wave oscillations is called the *Barkhausen criterion*. In this section a number of commonly used sine-wave oscillators are examined.

**PHASE-SHIFT OSCILLATOR** An example of a phase-shift oscillator using a FET is illustrated in Fig. 12.42. Biased for class A operation, the output of the FET is connected to three cascaded RC sections. In turn, the output of the three RC sections, which serves as the feedback network, is connected to the input (gate) of the amplifier.

The amplifier and the RC network each contributes  $180^\circ$  phase shift at the frequency of oscillation. Total phase shift, therefore, is  $180 + 180 = 360^\circ$ . This ensures that positive feedback exists and, if  $A_v$  and  $B$  are chosen properly, the Barkhausen criterion is satisfied. The magnitude of the amplifier gain must be at least 29 and the frequency of oscillation  $f$  in hertz is

$$f = \frac{0.159}{(\sqrt{6RC})} \quad (12.45)$$



**Fig. 12.42** An example of a phase-shift oscillator.



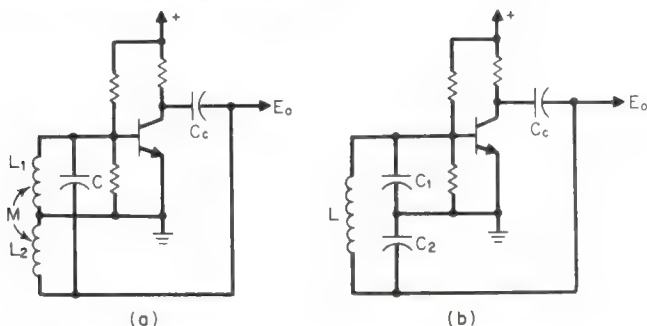


Fig. 12.43 Examples of tuned-circuit oscillators: (a) Hartley. (b) Colpitts.

**TUNED-CIRCUIT OSCILLATORS** Two examples of tuned-circuit, or *resonant circuit*, oscillators using a BJT are illustrated in Fig. 12.43. In the *Hartley* oscillator of Fig. 12.43a, appropriate phase shift to satisfy the Barkhausen criterion is obtained by the mutual coupling  $M$  between inductors  $L_1$  and  $L_2$ . The output of the amplifier is coupled to the input side by coupling capacitor  $C_c$ . Assuming small mutual coupling, an approximate expression for frequency of oscillation is

$$f = \frac{0.159}{\sqrt{C(L_1 + L_2)}} \quad (12.46)$$

In the *Colpitts* oscillator of Fig. 12-43b, suitable phase shift is obtained by capacitors  $C_1$ ,  $C_2$ , and inductor  $L$ . As in the Hartley oscillator,  $C_c$  couples the output voltage of the BJT to the input side. Because capacitor  $C_1$  is across the base-emitter junction of the transistor, the input capacitance  $C_i$  of the transistor should be included. Letting  $C'_1 = C_1 + C_i$ , the frequency of oscillation is given by

$$f = \frac{0.159}{\sqrt{LC'_1C_2/(C'_1 + C_2)}} \quad (12.47)$$

**FREQUENCY STABILITY** The frequency of oscillation tends to vary, or *drift*, with variations in temperature or aging. *Frequency stability* refers to the ability of an oscillator to maintain its frequency constant over long periods of time. To ensure frequency stability, various methods are employed. One method uses a capacitor with a negative-temperature coefficient (see Chaps. 2 and 10) to cancel the positive temperature coefficient of the inductor. Another approach is to use a very high- $Q$  tuned circuit. An example of a high- $Q$  circuit is the quartz crystal, which is the basis of the crystal oscillator.

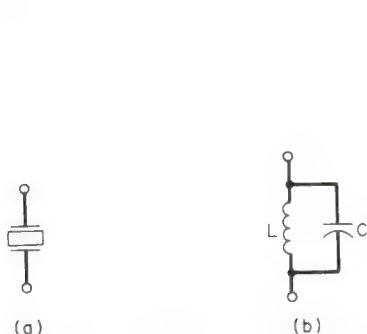


Fig. 12.44 Quartz crystal: (a) Electrical symbol. (b) Approximate model.

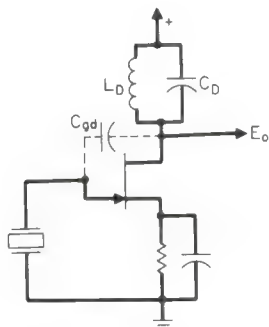


Fig. 12.45 An example of a crystal oscillator.

**CRYSTAL OSCILLATOR** Quartz crystal has the property of becoming mechanically deformed when a voltage is impressed across its faces. Or, if mechanically strained, it produces a voltage. This phenomenon is referred to as the *piezoelectric effect*. The electrical symbol and approximate model of a quartz crystal are shown in Fig. 12.44. An examination of the model shows that the quartz crystal acts like a parallel  $LC$  tuned circuit. Its  $Q$  may be as high as 10 000, thereby ensuring excellent frequency stability.

An example of a FET crystal oscillator is illustrated in Fig. 12.45. The crystal is connected to the input circuit of the amplifier, and an  $L_D C_D$  circuit tuned to the frequency of the crystal is in the output circuit. Suitable coupling to satisfy the Barkhausen criterion is obtained through the internal drain-gate capacitance  $C_{gd}$ , and stray capacitance. The frequency of oscillation is determined by the crystal. If, for example, a 2.5-MHz oscillator is needed, a 2.5-MHz crystal is inserted in the circuit.



# Chapter 13

## Operational Amplifiers

### 13.1 INTRODUCTION

The phenomenal development of monolithic IC technology has made the operational amplifier (*op amp*) perhaps the most versatile component in electronics. With proper selection of feedback elements, the op amp may be used as a precision gain-voltage amplifier, buffer, summer, current source, converter, oscillator, and in many more applications. Its field is generally limited by the user's ingenuity and imagination.

Actually, the operational amplifier is not new. It was, and still is, used in analog computers. (An analog computer is employed to simulate physical systems.) Because of IC technology, it is possible to obtain an op amp, containing a dozen transistors, in a TO-5 package, which is used for packaging a single discrete transistor. Also, the cost of some IC operational amplifiers are less than what one paid for a single transistor just a few years ago.

### 13.2 THE BASIC OP AMP

In this section we treat the op amp as a *black box* with ideal characteristics. By considering the op amp as an ideal component, we can easily derive results of its application in different circuits. Fortunately, the derivations also hold, with an error of a few percent, for the majority of commercially available op amps.

The basic op amp is represented by the symbol of Fig. 13.1. The input denoted by the negative sign is referred to as the *inverting input*; the plus sign denotes the *non-inverting input*. For simplicity, ground, power-supply, and frequency-compensation (to be considered later) connections are omitted from the figure.

If a signal is impressed across the inverting terminal with respect to ground, the output signal is  $180^\circ$  out of phase with respect to the input signal, as in the common-emitter amplifier. For a signal applied across the noninverting terminal with respect to ground, the output is in phase with the input. If signals are impressed across both the inverting and noninverting terminals with respect to ground, the output is proportional to the difference of the input signals.

**IDEAL CHARACTERISTICS** The attributes of the ideal op amp, which many physical op amps approach, are:

1. Voltage gain is infinite.
2. Input impedance is infinite.
3. Output impedance is zero.
4. The frequency response is flat (constant) for all frequencies.
5. The output signal is zero if the input signal is zero.

## 13-2 Operational Amplifiers

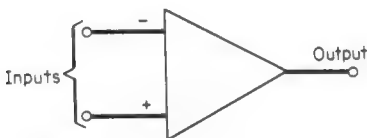


Fig. 13.1 Symbol for the basic op amp.

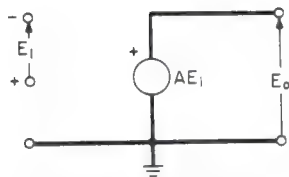


Fig. 13.2 Model of an ideal op amp.

Based on these attributes, a model of the ideal op amp is drawn in Fig. 13.2. The output voltage  $E_o$  equals the product of the gain  $A$  and the net input signal across the inverting and noninverting terminals  $E_i$ .

**FEEDBACK** Physical op amps can exhibit voltage gains as high as  $10^6$ . If the input voltage is finite, but close to zero, the output voltage of such a high-gain amplifier would soar to infinity. Actually, the output would be limited by the power-supply voltages. Under this condition, the amplifier is said to be *saturated*, and useless. To make the op amp useful and versatile, *feedback* is introduced around the amplifier.

Consider the operational circuit, called an *inverting amplifier*, of Fig. 13.3. The noninverting terminal is returned to ground and a feedback resistor  $R_F$  is connected between the inverting and output terminals. A resistor  $R_1$  is in series with the signal source  $E_s$  and the inverting terminal.

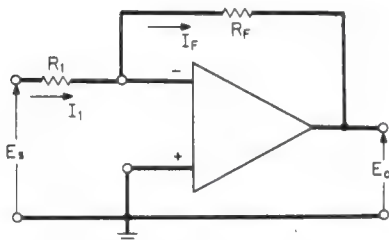


Fig. 13.3 An inverting amplifier.

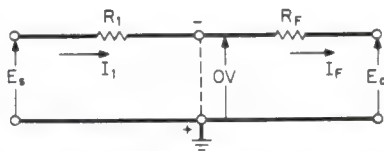


Fig. 13.4 The concept of the virtual ground.

As mentioned earlier, the ideal op amp exhibits infinite input impedance and voltage gain. The latter property implies that the voltage across the inverting and noninverting terminals is zero. Because the input impedance is infinite, no current flows into the inverting terminal. This condition is referred to as a *virtual ground* (Fig. 13.4). From the figure it is seen that  $I_1$ , the current flowing in  $R_1$ , is equal to  $I_F$ , the current flowing in  $R_F$ . Current  $I_1 = (E_s - 0)/R_1$  and  $I_F = (0 - E_o)/R_F$ . Hence,

$$\frac{E_s}{R_1} = \frac{-E_o}{R_F}$$

Solving for the voltage gain with feedback,  $A_f = E_o/E_s$ , we have

$$A_f = -\frac{R_F}{R_1} \quad (13.1)$$

What is striking and significant about Eq. (13.1) is that the voltage gain is only a function of resistors  $R_F$ ,  $R_1$ , and is independent of the amplifier. For example, if  $R_F = 100$  and  $R_1 = 1 \text{ k}\Omega$ ,  $A_f = -100$ . If  $R_F$  and  $R_1$  are precision resistors, the gain of  $-100$  will hold, regardless of variations in amplifier characteristics, power-supply voltage, etc. This behavior is typical of high-gain commercial op amps with suitable feedback.

### 13.3 APPLICATIONS

The inverting amplifier has already been analyzed. In this section, examples of other commonly used applications of the op amp are considered.

**NONINVERTING AMPLIFIER** The noninverting amplifier is shown in Fig. 13.5. In this circuit, the signal is fed directly to the noninverting terminal of the op amp. Resistors  $R_F$  and  $R_2$  comprise the feedback network. Because of the virtual ground, the voltage at the inverting and noninverting terminals is equal to  $E_s$ . Owing to the

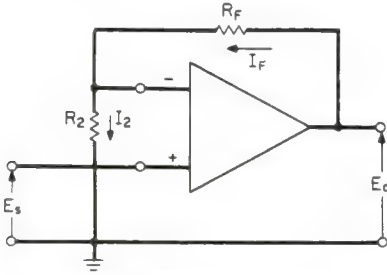


Fig. 13.5 A noninverting amplifier.

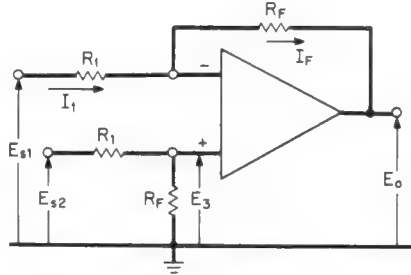


Fig. 13.6 A difference amplifier.

input impedance being infinite,  $I_F = I_2$ . Current  $I_F = (E_o - E_s)/R_F$  and current  $I_2 = E_s/R_2$ . Hence,

$$\frac{E_o - E_s}{R_F} = \frac{E_s}{R_2}$$

Solving for  $A_f = E_o/E_s$ ,

$$A_f = \frac{(R_2 + R_F)}{R_2} = 1 + \frac{R_F}{R_2} \quad (13.2)$$

For the noninverting amplifier, the output and input voltages are in phase; for the inverting amplifier, they are  $180^\circ$  out of phase.

**DIFFERENCE AMPLIFIER** The difference amplifier of Fig. 13.6 provides an output that is proportional to the difference of the input signals,  $E_{s1} - E_{s2}$ . Using the concept of the virtual ground, the voltage at the inverting and noninverting terminals is equal to voltage  $E_3$ . Current  $I_1 = (E_{s1} - E_3)/R_1$  and  $I_F = (E_3 - E_o)/R_F$ . Because  $I_1 = I_F$ ,

$$\frac{E_{s1} - E_3}{R_1} = \frac{E_3 - E_o}{R_F}$$

By voltage division,  $E_3 = E_{s2}R_F/(R_1 + R_F)$ . Substitution of this expression in the preceding equation yields

$$\frac{E_{s1} - E_{s2}R_F/(R_1 + R_F)}{R_1} = \frac{E_{s2}R_F/(R_1 + R_F) - E_o}{R_F}$$

Solution for the last equation for the output voltage  $E_o$  gives

$$E_o = -\frac{R_F}{R_1} (E_{s1} - E_{s2}) \quad (13.3)$$

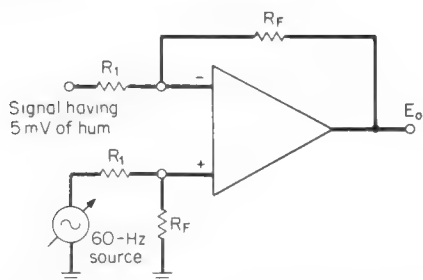
Thus the output voltage is proportional to the difference in the input voltages.

**example 13.1** An input signal feeding an inverting amplifier has 5 mV of 60-Hz hum. How can a difference amplifier be used to cancel the hum?

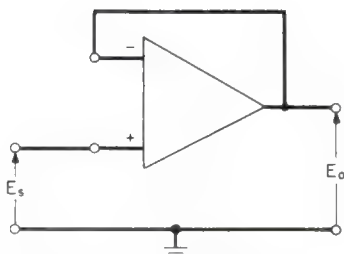
**solution** The circuit using a difference amplifier is shown in Fig. 13.7. The variable-voltage 60-Hz source is adjusted until no 60-Hz hum appears in the output  $E_o$ . This condition is monitored on a scope connected across the output terminal and ground of the op amp.



### 13-4 Operational Amplifiers



**Fig. 13.7** Cancellation of 60-Hz hum present in a signal.



**Fig. 13.8** A voltage follower.

**VOLTAGE FOLLOWER** An op amp circuit which has unity voltage gain, called a voltage follower, is shown in Fig. 13.8. Voltage followers, like emitter and source followers, are used for matching a high-impedance source to a low-impedance load.

The output terminal of the op amp is connected directly to the inverting terminal; the signal is impressed between the noninverting terminal and ground. Because of the virtual ground, the voltage at the inverting terminal is equal to the voltage at the noninverting terminal  $E_s$ . Since the output is connected to the inverting terminal,  $E_o = E_s$ , and

$$A_f = \frac{E_o}{E_s} = 1 \quad (13.4)$$

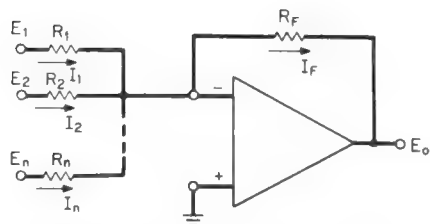
**SUMMING AMPLIFIER** The summing amplifier of Fig. 13.9 provides an output voltage proportional to the sum of the input voltages,  $E_1, E_2, \dots, E_n$ . If resistors  $R_1 = R_2 = \dots = R_F$ , then the output is equal to the negative sum of the inputs. If the resistors are unequal, then each input is summed with a *scale factor*.

Owing to the virtual ground,  $I_1 = E_1/R_1, I_2 = E_2/R_2, \dots, I_n = E_n/R_n$  and  $I_F = -E_o/R_F$ . Applying Kirchhoff's current law (KCL), the algebraic sum of the currents at the inverting terminal equals zero. Hence,

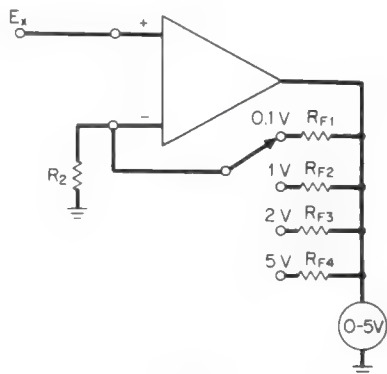
$$\frac{E_1}{R_1} + \frac{E_2}{R_2} + \dots + \frac{E_n}{R_n} = \frac{-E_o}{R_F}$$

Solving for  $E_o$ , we obtain

$$E_o = -\left(\frac{R_F}{R_1}E_1 + \frac{R_F}{R_2}E_2 + \dots + \frac{R_F}{R_n}E_n\right) \quad (13.5)$$



**Fig. 13.9** A summing amplifier.



**Fig. 13.10** An example of a multirange dc voltmeter. (See Example 13.3.)

**example 13.2** Three input voltages are to be summed in the summing amplifier of Fig. 13.9. The feedback resistor  $R_f = 100 \text{ k}\Omega$ ,  $E_1 = 0.1 \text{ V}$ ,  $E_2 = 0.2 \text{ V}$ , and  $E_3 = 0.05 \text{ V}$ . Determine  $E_o$  if (a)  $R_1 = R_2 = R_3 = 100 \text{ k}\Omega$  and (b)  $R_1 = 20$ ,  $R_2 = 50$ ,  $R_3 = 10 \text{ k}\Omega$ .

**solution** (a) For equal resistors, Eq. (13.5) reduces to  $E_o = -(E_1 + E_2 + E_3) = -(0.1 + 0.2 + 0.05) = -0.35 \text{ V}$ .

(b) For unequal input resistors, each input voltage, from Eq. (13.5), is multiplied by a scale factor:  $R_f/R_1$ ,  $R_f/R_2$ , and  $R_f/R_3$ . Hence,

$$\begin{aligned} E_o &= -\left(\frac{100}{20} \times 0.1 + \frac{100}{50} \times 0.2 + \frac{100}{10} \times 0.05\right) \\ &= -(5 \times 0.1 + 2 \times 0.2 + 10 \times 0.05) \\ &= -(0.5 + 0.4 + 0.5) = -1.4 \text{ V} \end{aligned}$$

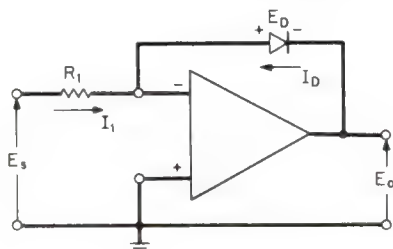
**example 13.3** A precision multirange dc voltmeter is to be constructed. The voltage ranges desired are zero to 0.1, 1, 2, and 5 V full scale. Using an op amp and a 0- to 5-V dc meter, design the circuit.

**solution** A design of such a meter is shown in Fig. 13.10. The unknown voltage to be measured,  $E_x$ , is impressed across the noninverting terminal and ground. To realize the four voltage ranges, a four-position switch and four resistors,  $R_{F1}$ ,  $\dots$ ,  $R_{F4}$ , are connected in the feedback path. Resistor  $R_2$ , connected to the inverting terminal and the switch, completes the feedback circuit. A reasonable value for  $R_2$  is  $5 \text{ k}\Omega$ .

When  $E_x = 0.1 \text{ V}$ , the 0- to 5-V meter must read 0.1 V full scale. The gain of the noninverting amplifier is, therefore, equal to  $5/0.1 = 50$ . From Eq. (13.2),  $A_f = 1 + R_f/R_2$ ; hence,  $50 = 1 + R_{F1}/5$ . Solving,  $R_{F1} = (50 - 1) \times 5 = 245 \text{ k}\Omega$ .

Following the same procedure, for the 0- to 1-V scale,  $A_f = 5/1 = 5$ . Therefore,  $R_{F2} = (5 - 1) \times 5 = 20 \text{ k}\Omega$ . For the 0- to 2-V scale,  $A_f = 5/2 = 2.5$ ;  $R_{F3} = (2.5 - 1) \times 5 = 7.5 \text{ k}\Omega$ . Finally, for the 0- to 5-V scale,  $A_f = 5/5 = 1$ ;  $R_{F4} = (1 - 1) \times 5 = 0$ .

**LOGARITHMIC AMPLIFIER** This circuit provides an output voltage that is proportional to the logarithm of the input voltage. A simple version of a logarithmic amplifier



**Fig. 13.11** An elementary logarithmic amplifier.

is obtained by replacing the feedback resistor  $R_f$  in the inverting amplifier of Fig. 13.3 by a diode (Fig. 13.11). The diode current  $I_D$  may be approximated by

$$I_D = I_o e^{+E_D/k} \quad (13.6)$$

where  $I_o$  is the diode reverse saturation current

$E_D$  is the voltage across the diode

$k$  is a term that includes physical constants and temperature

Because of the virtual ground,  $I_1 = E_s/R_1 = I_D$ . Also,  $E_D = 0 - E_o = -E_o \text{ V}$ . Making these substitutions in Eq. (13.6), we obtain

$$\frac{E_s}{R_1} = I_o e^{-E_o/k} \quad (13.7)$$

Taking the natural logarithm ( $\ln$ ) of Eq. (13.7) yields

$$\ln \frac{E_s}{R_1} = \ln I_o - \frac{E_o}{k}$$

Therefore,

$$E_o = -k \left( \ln \frac{E_s}{R_1} - \ln I_o \right) \quad (13.8)$$

### 13-6 Operational Amplifiers

Equation (13.8) demonstrates that the output voltage is indeed proportional to the natural log of the input voltage.

**CURRENT-TO-VOLTAGE CONVERTER** Devices, such as photomultipliers and photocells, act like an ideal current source. The output current of these devices is essentially constant, independent of the load. The current-to-voltage converter provides an output voltage which is proportional to the current.

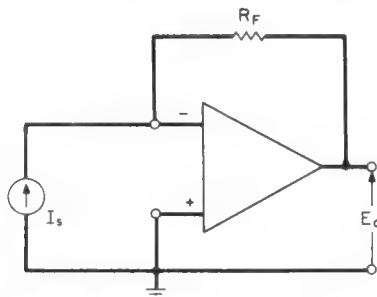


Fig. 13.12 A current-to-voltage converter.

An example of such a converter is illustrated in Fig. 13.12. The device is represented by the ideal current source  $I_s$ . Because of the virtual ground,

$$I_s = \frac{-E_o}{R_F}$$

or

$$E_o = -I_s R_F \quad (13.9)$$

The output voltage is, therefore, proportional to the input current.

### 13.4 PRACTICAL CONSIDERATIONS

To this point we represented the op amp as having only three terminals: the inverting and noninverting input terminals and an output terminal. In addition to these, there are terminals for ground, power supply, and compensating network connections (Fig. 13.13).

Generally negative ( $-E$ ) and positive ( $+E$ ) power supplies are required. There are at least two terminals to which are connected a compensating network. Compensation is required to ensure stable amplifier operation. The compensating network usually consists of a small capacitor in series with a resistor.

**FREQUENCY RESPONSE AND COMPENSATION** A frequency response plot of an *uncompensated* operational amplifier may take the form shown in Fig. 13.14 (solid lines). Along the vertical axis is plotted the magnitude of the voltage gain without feedback, or the *open-loop gain*, in decibels (dB). A voltage gain of 100 dB corre-

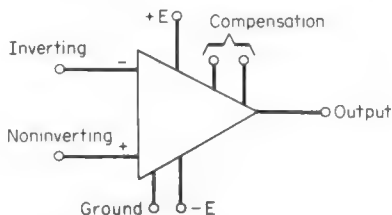


Fig. 13.13 Symbol for a physical op amp.

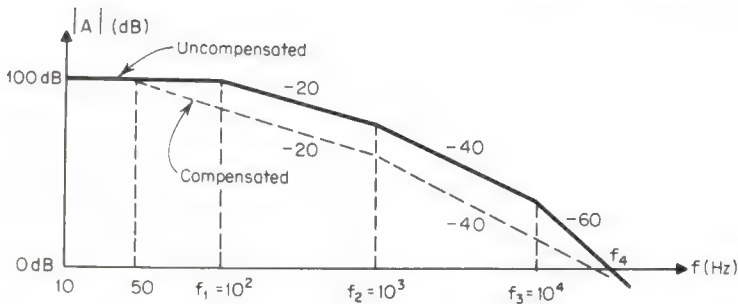


Fig. 13.14 Bode response plots of an uncompensated and a compensated op amp.

sponds to an amplifier gain of 100 000. Frequency is plotted to a logarithmic scale along the horizontal axis. This permits a large range of frequencies to be plotted in a compact manner.

The actual curve is approximated by straight-line segments. Such a representation is referred to as a *Bode*, or an *asymptotic*, plot. The Bode plot is easy to draw and provides useful information.

In Fig. 13.14 frequency  $f_1 = 100$  Hz is called a *break frequency*, where the gain is actually 3 dB less than at lower frequencies. For frequencies between  $f_1$  and  $f_2$ , the gain decreases, or *rolls off*, -20 dB for a decade of frequency. This means that for a 10-to-1 increase in frequency, the voltage gain falls by 20 dB. An *equivalent expression*, -6 dB/octave, means that as the frequency is *doubled*, the gain decreases by 6 dB. At frequency  $f_2$  the gain begins to roll off at -40 dB/decade (-12 dB/octave), and at  $f_3$  it rolls off at -60 dB/decade (-18 dB/octave). Frequency  $f_4$ , which intersects the frequency axis at unity voltage gain (0 dB), is the *crossover frequency*.

An amplifier that exhibits an open-loop response of -60 dB/decade (or -18 dB/octave) at the crossover frequency is *unstable*. It can break into oscillation and generate sine waves. By the proper selection of a compensating network, the amplifier is made stable. For this condition, indicated by the dashed lines in Fig. 13.14, the slope is -40 dB/decade (-12 dB/octave) at the crossover frequency.

The specific compensation network to use is recommended by the manufacturer of the op amp. It may, for example, be a 2200-pF capacitor connected in series with a 10-k $\Omega$  resistor.

From Fig. 13.14 the dominant frequency of the compensated amplifier is only 50 Hz. It might appear, therefore, that the op amp is really not too useful. Because of the use of feedback, which increases the bandwidth of an amplifier, this is not true.

Consider an op amp with an open-loop gain of 100 000 (100 dB). If, for an inverting amplifier,  $A_f = 100$  (which equals  $20 \log 100 = 40$  dB), the difference between the open-loop gain and the gain with feedback, referred to as the *closed-loop gain*, is  $100 - 40 = 60$  dB at low frequencies. This is illustrated in Fig. 13.15. Note that the

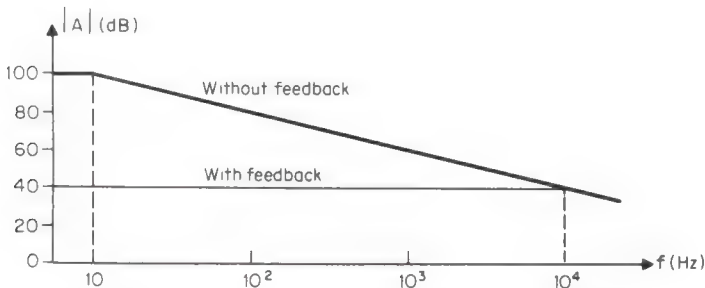


Fig. 13.15 Feedback increases the bandwidth of an amplifier.

### 13-8 Operational Amplifiers

bandwidth is much greater than 10 Hz; in this example it is 10 kHz. For frequencies greater than 10 kHz, the open- and closed-loop gains are equal.

**OFFSET** For a zero input, an ideal op amp has exactly zero output. In physical op amps, because of variations in transistor characteristics, there is a nonzero output for a zero signal. To compensate for this, a dc voltage is applied to the amplifier such that, for a zero input signal, the output is also zero. Before we see how this is accomplished, some terms used in the literature and on data sheets are defined.

**Input offset current** The input offset current  $I_{io}$  is the difference in dc bias currents in an op amp.

**Input offset voltage** The input offset voltage  $E_{io}$  is the voltage impressed across the input of an op amp to ensure that the output is zero when the input signal is also zero.

**Input offset current drift** The input offset current drift is the ratio of input offset current change  $\Delta I_{io}$  to a change in temperature  $\Delta T$ . It is expressed by  $\Delta I_{io}/\Delta T$ .

**Input offset voltage drift** The input offset voltage drift is the ratio of input offset voltage change  $\Delta E_{io}$  to a change in temperature  $\Delta T$ . It is expressed by  $\Delta E_{io}/\Delta T$ .

**OFFSET CORRECTION** A circuit for offset correction in an inverting amplifier is illustrated in Fig. 13.16. Connected to the noninverting input is a voltage divider

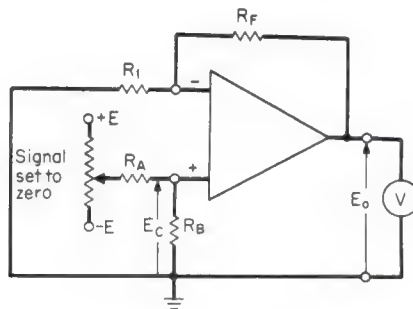


Fig. 13.16 Offset correction of an inverting amplifier.

consisting of resistors  $R_A$  and  $R_B$ . One end of resistor  $R_A$  is connected to the wiper arm of the potentiometer (pot). A positive voltage ( $+E$ ) is connected to one end of the pot, and the other end is connected to a negative ( $-E$ ) voltage. This arrangement makes available both positive and negative voltages for offset correction. The value of the pot is in the order of 20 k $\Omega$ .

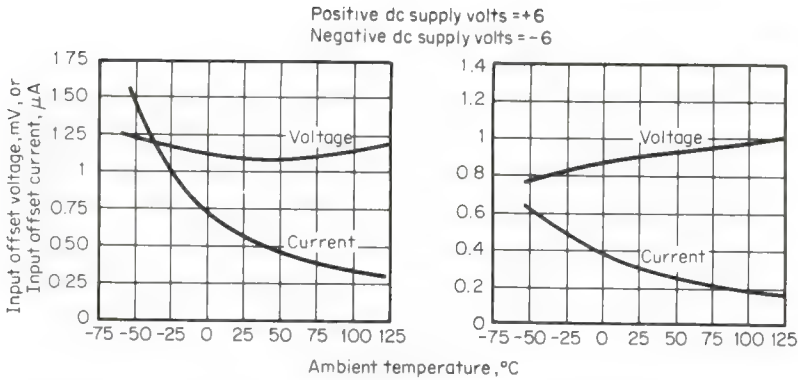
Offset correction is done as follows: With the signal set to zero, a sensitive dc voltmeter is connected across the output to monitor output voltage  $E_o$ . The pot is then adjusted until  $E_o = 0$ . It must be emphasized that, because offset voltage and current drift with temperature, the correction obtained is only valid for a *given operating temperature*. Variations in offset current and voltage with temperature are illustrated in Fig. 13.17.

**example 13.4** Referring to Fig. 13.16, assume that  $\pm E = \pm 20$  V,  $R_A = 250$  k $\Omega$ , and  $R_B = 100$   $\Omega$ . Determine the maximum value of the offset correction voltage  $E_C$ .

**solution** By voltage division,  $E_C = \pm 20 \times 100/(250\,000 + 100) \approx \pm 2000/250\,000 = \pm 8$  mV.

A circuit for correcting offset in a noninverting amplifier is given in Fig. 13.18. As for the inverting amplifier, with a dc voltmeter connected across the output, the pot is adjusted until  $E_o = 0$ .

**SLEWING RATE** An important characteristic of an op amp is how well it follows a rapidly changing waveform. Assume that the input to a noninverting amplifier, with a

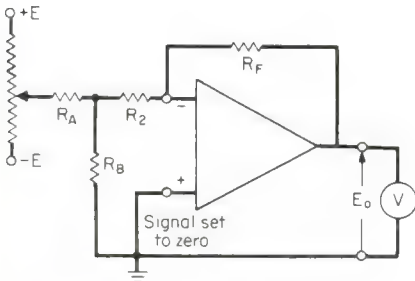


**Fig. 13.17** Examples of how input offset current and voltage vary with temperature. (Courtesy RCA.)

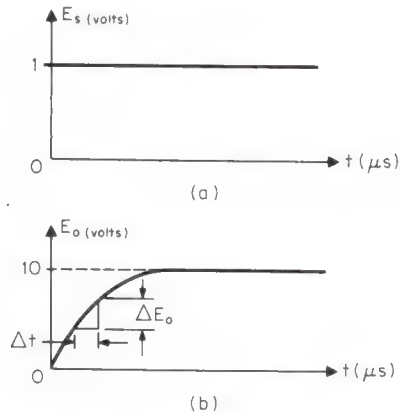
gain of 10, is the step input signal of Fig. 13.19a. The output is shown in Fig. 13.19b. Note that the output voltage of 10 V is not reached instantaneously. The rate of change of the rising voltage  $\Delta E_o$  with respect to time  $\Delta t$  is the *slewing rate*:

$$\text{Slewing rate} = \frac{\Delta E_o}{\Delta t} \quad (13.10)$$

The unit of the slewing rate is usually expressed as volts per microsecond. A high slewing rate indicates an op amp with a good response to rapidly changing waveforms.



**Fig. 13.18** Offset correction of a noninverting amplifier.



**Fig. 13.19** Slewing rate of an op amp: (a) Step input to a noninverting amplifier. (b) Output of amplifier.

### 13.5 OP AMP CIRCUITRY

An operational amplifier contains a number of stages, each of which performs a specific function. This is illustrated in the block diagram (Fig. 13.20) of a typical op amp. Each of the circuits will be first briefly described and then examined in detail.

The differential amplifier in the first stage provides the inverting and noninverting inputs, gain, and is very stable in operation. The constant-current sink may be thought of as simulating a high emitter resistance with good efficiency. For additional gain, a second amplifier stage is required. Often, the second stage is another differential amplifier.



### 13-10 Operational Amplifiers

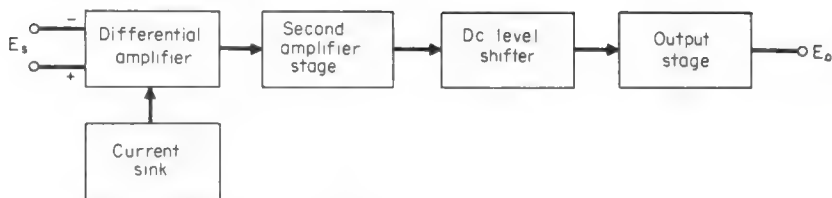


Fig. 13.20 General block diagram of an operational amplifier.

Because the circuits in an op amp are direct-coupled, as the signal progresses from the input to the output there is a buildup in dc voltage. The dc level shifter cancels out the built-up dc voltage. This ensures that, neglecting offset, the output is zero for zero input signal.

The output stage may be an emitter follower. For high-level signals, however, a push-pull type of output stage is used.

**DIFFERENTIAL AMPLIFIER** A basic differential amplifier is illustrated in Fig. 13.21a. A striking feature of the circuit is its *symmetry*. Because IC processing yields devices that are closely matched, the differential amplifier may be viewed as a Wheatstone bridge shown in Fig. 13.21b. Two power supplies are generally required for its operation:  $+E_{CC}$  and  $-E_{EE}$  volts. Input signals  $E_1$  and  $E_2$ , or either one alone, are applied to the transistors,  $Q_1$  and  $Q_2$ , and  $Q_2$ , respectively.

Assume that the input signals are equal:  $E_1 = E_2$ . For this condition, the output voltages are also equal ( $E_{o1} = E_{o2}$ ) and, viewed as a bridge, the bridge is said to be balanced. If input  $E_1$  is greater than  $E_2$ , more collector current flows in  $Q_1$  and less collector current in  $Q_2$ . Hence,  $E_{o1} < E_{o2}$ . On the other hand, if  $E_2 > E_1$ , more collector current now flows in  $Q_2$  than in  $Q_1$ , and  $E_{o1} > E_{o2}$ .

Practical signals may be thought of as containing two components: the *common-mode* ( $E_c$ ) and *difference-mode* ( $E_d$ ) signals. The common-mode signal is *common to both inputs*. Examples of this kind of signal include noise, hum from a power supply, and effects of temperature change. The difference-mode signal is the signal that is to be amplified. For an ideal differential amplifier, therefore, the gain to the common-mode signal would be zero, and the gain to the difference-mode signal would be very high.

Neglecting  $h_{re}$  and  $h_{oe}$  of the transistors, the common-mode signal gain  $A_c$  is given by

$$A_c \approx \frac{-R_C}{2R_E} \quad (13.11)$$

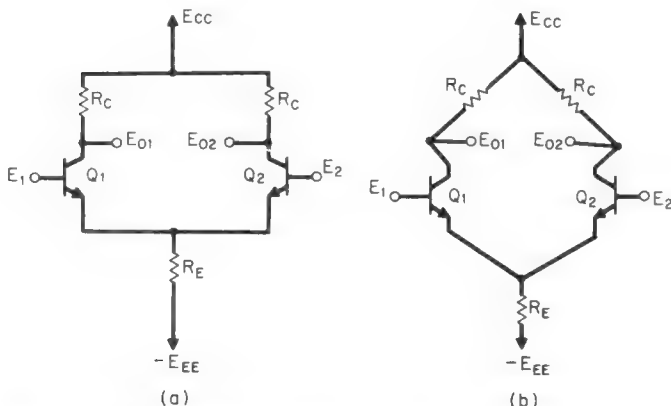


Fig. 13.21 Basic differential amplifier: (a) Circuit. (b) Amplifier viewed as a bridge.

and the difference-mode signal gain  $A_d$  is

$$A_d \approx \frac{-h_{fe}R_C}{2h_{ie}} \quad (13.12)$$

From Eq. (13.11), for a given value of  $R_C$ , a small common-mode gain requires that  $R_E$  should be large. For very practical reasons,  $R_E$  cannot be made too large. For one, the practical limit of diffused resistors is 30 k $\Omega$ . Furthermore, for a large value of  $R_E$ , excessive power is wasted as heat. For these reasons a current sink, which simulates electronically a large value of resistance, is used in place of  $R_E$ . (The current sink is treated later in this chapter.)

Referring to Eq. (13.12), for a high value of difference-mode gain,  $h_{fe}$  and  $R_C$  should be large and  $h_{ie}$  low. High-gain transistors can be fabricated by making the bases exceedingly thin. One such device, called the *super beta transistor*, has an  $h_{fe}$  of a few thousand. For practical reasons, the maximum value of  $R_C$  is 30 k $\Omega$ . The value of  $h_{ie}$  varies inversely with collector bias current. Because the transistors in a differential amplifier are biased at a low quiescent current, a typical value of  $h_{ie}$  is a few kilohms.

A figure of merit for differential (and operational) amplifiers is the *common-mode rejection ratio*, CMRR. It is defined as the ratio of the difference-mode gain to the common-mode gain:

$$\text{CMRR} = \frac{A_d}{A_c} \quad (13.13)$$

The CMRR is generally expressed in decibels. The greater the value of CMRR, the better is the differential (or operational) amplifier.

**example 13.5** Referring to the differential amplifier of Fig. 13.21a, assume that  $R_C = R_E = 10$  k $\Omega$ ,  $h_{fe} = 100$ , and  $h_{ie} = 4$  k $\Omega$ ;  $h_{re}$  and  $h_{oe}$  are negligible. Determine (a)  $A_c$ , (b)  $A_d$ , and (c) CMRR.

**solution** (a) From Eq. (13.11),  $A_c \approx -R_C/2R_E = -10/(2 \times 10) = -0.5$ .

(b) From Eq. (13.12),  $A_d \approx -h_{fe}R_C/2h_{ie} = -100 \times 10/(2 \times 4) = -125$ .

(c)  $\text{CMRR} = A_d/A_c = -125/(-0.5) = 250$ . In decibels,  $\text{CMRR} = 20 \log (250) \approx 48$  dB.

**CURRENT SINK** In the previous section it was seen that for a low value of common-mode gain, resistor  $R_E$  should be large. An efficient method for simulating a high-value resistance is by use of a *current sink*. An example of such a circuit connected to the emitters of a differential amplifier is shown in Fig. 13.22a.

Transistor  $Q_3$  acts like a constant-current source. Resistors  $R_1$  and  $R_2$  bias the transistor and  $R_3$  stabilizes its quiescent operating point. The circuit of Fig. 13.22a simulates a resistance of a few hundred kilohms. A general symbol for a current sink, often found in the literature, is illustrated in Fig. 13.22b.

An example of a commercially integrated differential amplifier, the Fairchild  $\mu\text{A}730$ , is shown in Fig. 13.23. Transistors  $Q_1$  and  $Q_2$  comprise the differential amplifier;  $Q_3$  is the current sink. For low output impedance,  $Q_3$  and  $Q_4$  are connected as emitter followers.

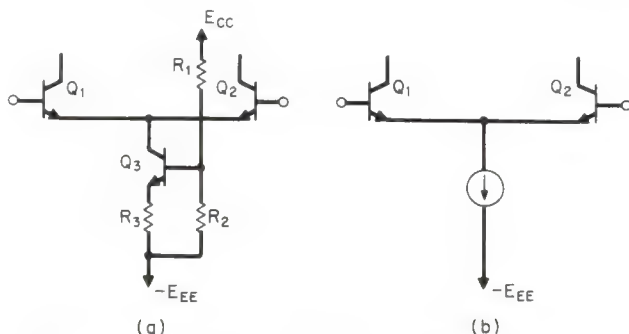


Fig. 13.22 Constant-current sink: (a) Circuit. (b) Symbol.

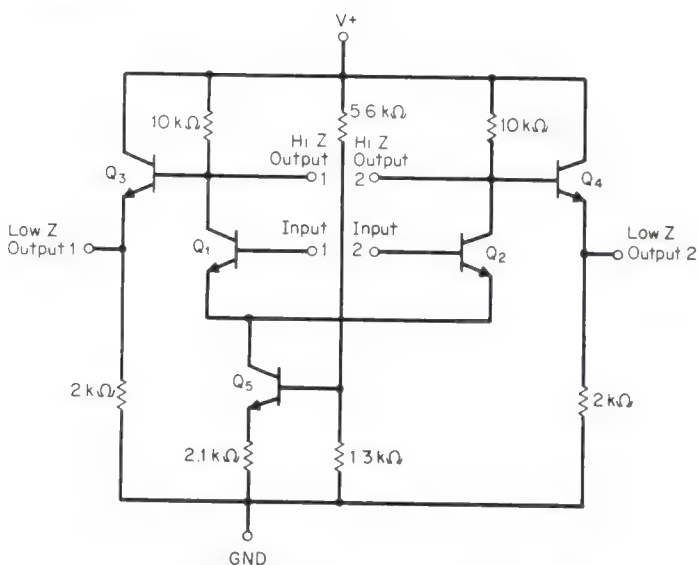


Fig. 13.23 Example of a commercial IC ( $\mu A730$ ) differential amplifier. (Courtesy Fairchild Semiconductor.)

**CASCADING STAGES** To achieve high gain, an additional amplifier stage is cascaded to the differential amplifier. An example is the cascaded pair of differential amplifiers shown in Fig. 13.24. Transistors  $Q_1$  and  $Q_2$  and the constant current sink comprise the first differential amplifier stage. The second differential-amplifier stage contains transistors  $Q_3$  and  $Q_4$ . Because only a single-ended output  $E_o$  from  $Q_4$  is required, no collector resistor is necessary for  $Q_3$ . Also, an emitter resistor  $R_E$ , instead of a current sink, is adequate for the second stage.

Another circuit for increasing gain is illustrated in Fig. 13.25. Transistors  $Q_1$  and  $Q_2$  constitute the differential amplifier;  $Q_3$  and  $Q_4$  provide increased gain and convert the double-ended output of the differential amplifier to a single-ended output. The bases of  $Q_3$  and  $Q_4$  are fed from a common voltage point  $P$  through equal resistors  $R_{C1}$ . Transistor  $Q_3$  inverts the output of  $Q_1$  and compares it with the output of  $Q_2$ . The full differential output of the differential amplifier, therefore, is impressed across  $Q_4$ . Transistor  $Q_4$  amplifies the differential signal and provides a single-ended output.

Another interesting feature of the circuit is that it is relatively insensitive to changes in the positive supply voltage  $E_{CC}$ . If, for example,  $E_{CC}$  should rise, the collector cur-

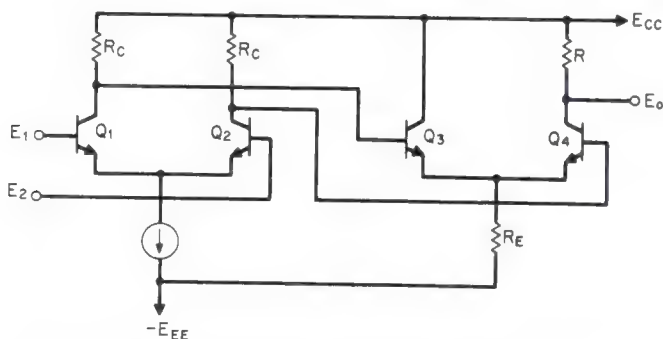
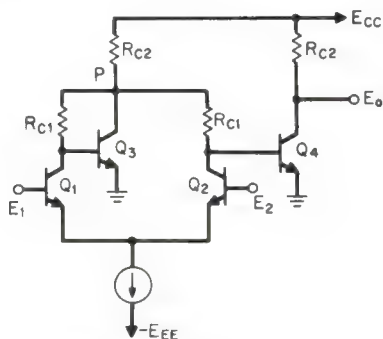


Fig. 13.24 A cascaded pair of differential amplifiers for increased gain.

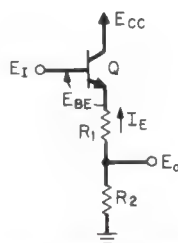
rents in  $Q_3$  and  $Q_4$  also rise. Consequently, the collector voltages of  $Q_3$  and  $Q_4$  tend to decrease. As their voltage falls, point  $P$  becomes less positive, thereby reducing the base current to  $Q_4$ . If the base current is reduced, its collector voltage increases. Thus we have a regulating (feedback) mechanism in operation that ensures the collector voltage of  $Q_4$  remaining essentially constant.

**DC LEVEL SHIFTER** In monolithic integrated circuits, coupling capacitors are not practical. Instead, direct coupling is used and, as the signal travels from the input to the output stage, there is a buildup in dc voltage. If this built-up voltage is not canceled, a nonzero output is obtained for zero signal.

A circuit for canceling the built-up dc voltage in an op amp is called a *dc level shifter*. There are many circuits for accomplishing this; a basic dc level shifter is shown in Fig. 13.26.



**Fig. 13.25** Transistors  $Q_3$  and  $Q_4$  provide additional gain and convert the double-ended output of the differential amplifier ( $Q_1$ ,  $Q_2$ ) into a single-ended output voltage,  $E_o$ .



**Fig. 13.26** A basic dc level shifter.

In this circuit, transistor  $Q$  is connected as an emitter follower. The emitter resistance consists of two resistors:  $R_1$  in series with  $R_2$ . The dc base-emitter voltage  $E_{BE}$  is nominally 0.7 V at room temperature. Let  $E_I$  be the dc voltage at the base of  $Q$  and  $E_O$  the output dc voltage. The shifted voltage is equal to  $E_I - E_O$  and is given by

$$E_I - E_O = E_{BE} + I_E R_1 \quad (13.14)$$

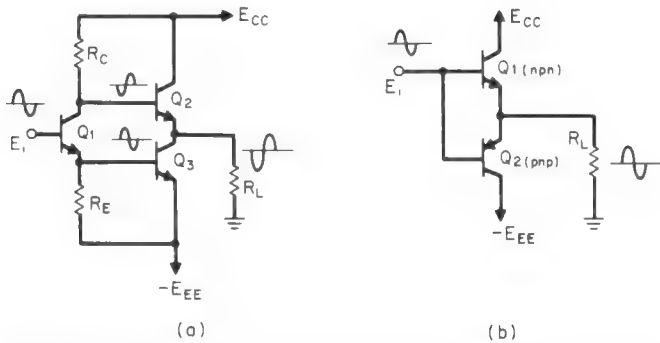
For example, if  $I_E = 2$  mA and  $R_1 = 600 \, \Omega$ , then the dc voltage shift is, by Eq. (13.14),  $0.7 + 2 \times 0.6 = 1.9$  V.

**OUTPUT STAGE** The output stage of an op amp must exhibit low output impedance and be capable of providing the required load power, and its output signal should be limited only by the supply voltages for a maximum input signal. For low output power, an emitter follower is often used as the output stage in an op amp. If the power requirements are large, however, excessive power is dissipated in the emitter resistor. For this case, the circuits of Fig. 13.27 are generally employed as the output stage of an op amp.

In the circuit of Fig. 13.27a, transistor  $Q_1$  operates as a *phase splitter*. If  $R_C = R_E$ , the signals are equal and 180 out of phase at the collector and in phase at the emitter with respect to input signal  $E_I$ . Transistors  $Q_2$  and  $Q_3$  operate as a push-pull amplifier, often referred to as a *totem-pole amplifier*.

Assume, for example, that the base voltage to  $Q_2$  is decreasing and the base voltage to  $Q_3$  is increasing. Transistor  $Q_3$ , therefore, conducts more current than  $Q_2$ . The difference in transistor currents is the load current, which flows in load resistor  $R_L$ . In many integrated circuits, emitter resistor  $R_E$  of  $Q_1$  is replaced by a diode.

The circuit of Fig. 13.27b, called a *complementary pair* output stage, eliminates the



**Fig. 13.27** Examples of output stages used in op amps: (a) Totem pole. (b) Complementary pair.

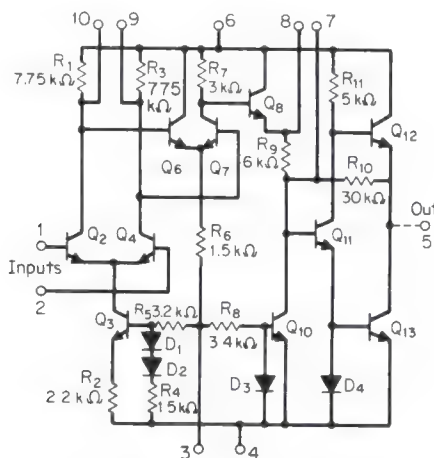
need for a phase splitter. Transistors  $Q_1$  and  $Q_2$  are of opposite types:  $Q_1$  is npn and  $Q_2$  is pnp. For efficient operation,  $Q_1$  and  $Q_2$  should be matched.

On the positive half-cycle of input signal  $E_i$ ,  $Q_1$  is conducting (because it is npn) and  $Q_2$  is nonconducting (because it is pnp). Current, therefore, flows in the positive direction through  $Q_1$  to load resistor  $R_L$ . On the negative half-cycle,  $Q_1$  is nonconducting and  $Q_2$  is conducting. Current flows in the negative direction through the pnp transistor  $Q_2$  to  $R_L$ , thereby completing the output waveform.

### 13.6 EXAMPLES OF PRACTICAL OP AMPS

Two examples of commercially available op amps will now be analyzed. The first is the Motorola MC-1530 op amp of Fig. 13.28. Transistors  $Q_2$  and  $Q_4$  comprise the first differential amplifier stage, which is cascaded to a second differential amplifier containing  $Q_6$  and  $Q_7$ . A current sink,  $Q_3$ , is used for the first stage and an emitter resistor,  $R_6$ , for the second stage.

Dc level shifting is achieved primarily by transistors  $Q_8$  and  $Q_{10}$ . A phase splitter,  $Q_{11}$ , drives the totem-pole amplifier containing transistors  $Q_{12}$  and  $Q_{13}$ . Note that a diode is used instead of an emitter resistor in the phase splitter.



**Fig. 13.28** Circuit of the Motorola MC-1530 op amp. (Courtesy Motorola Semiconductor Products, Inc.)

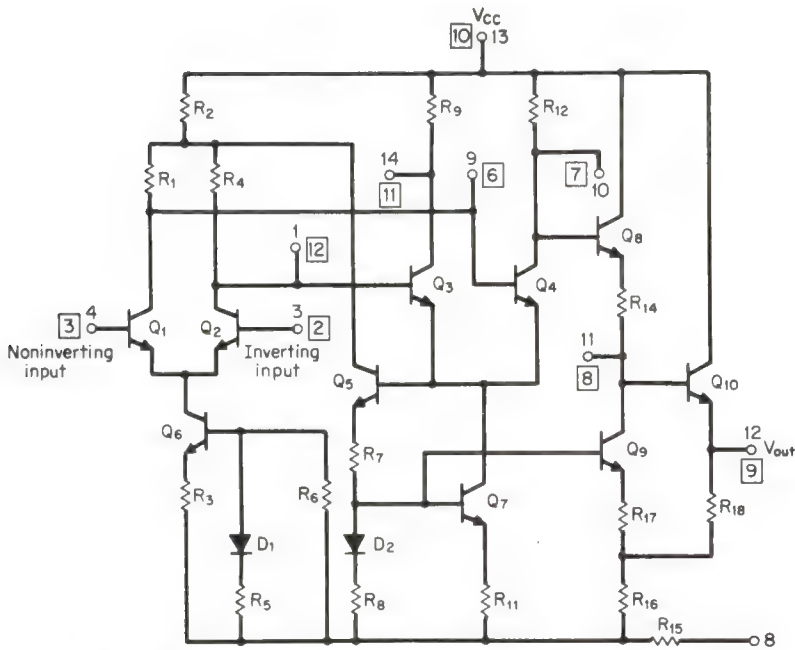


Fig. 13.29 Circuit diagram of the RCA CA 3008 op amp. (Courtesy RCA.)

As a second example, consider the RCA CA 3008 op amp of Fig. 13.29. Transistors  $Q_1$  and  $Q_2$  constitute the first differential amplifier stage which is coupled to a second differential amplifier ( $Q_3$ ,  $Q_4$ ). Current sinks are used for both stages ( $Q_6$ ,  $Q_7$ ), and diodes  $D_1$  and  $D_2$  provide thermal stabilization for their operating points.

Transistor  $Q_5$  compensates the effect of common-mode signals. Assume, for example, that the dc power-supply voltage decreases. As a result, the voltage at the emitters of  $Q_3$  and  $Q_4$  also decrease. Less collector current flows in  $Q_5$ , as well as in  $Q_7$  and  $Q_9$ . Because of the decrease in collector current of  $Q_7$ , which acts as a sink, less current flows in  $Q_3$  and  $Q_4$ . This results in an increase in their collector voltage, thereby canceling the decrease in the supply voltage.

Transistors  $Q_8$  and  $Q_9$  provide dc level shifting. The output stage is an emitter follower. Note that emitter resistor  $R_{18}$  of the follower is returned to the junction of resistors  $R_{16}$  and  $R_{17}$ , instead of to the negative supply voltage. This connection provides *positive feedback* and increases the gain of the emitter follower from less than 1 to about 1.5.

The current flowing in  $R_{18}$  is added to the current in  $R_{16}$ . The voltage drop across  $R_{16}$  is thereby increased. Because the power-supply voltages are fixed, the voltage across  $R_{14}$  is reduced. This results in the collector voltage of  $Q_9$ , which is connected to the base of the emitter follower, to increase. The greater input voltage to the emitter follower results in a higher output voltage of the op amp. Resistor  $R_{15}$  limits the signal load current in case output terminal 12 is short-circuited to terminal 8.

### 13.7 UNDERSTANDING THE DATA SHEET FOR AN OP AMP

In selecting an op amp for a particular application, the user must consult the data (specs) sheet provided by the manufacturer of the device. Representative data sheets for an op amp, the Fairchild  $\mu A709C$ , are reproduced in Fig. 13.30. Invariably, the data included in the sheets are:

**1. General description** A cogent statement that stresses special features and application areas of the op amp under consideration. For example, the  $\mu A709C$  features



OCTOBER 1967

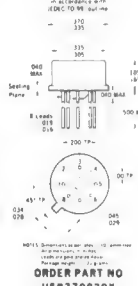
# **μA709C** HIGH PERFORMANCE OPERATIONAL AMPLIFIER FAIRCHILD LINEAR INTEGRATED CIRCUITS

**GENERAL DESCRIPTION** - The μA709C is a high-gain operational amplifier constructed on a single silicon chip using the Fairchild Planar epitaxial process. It features low offset, high input impedance, large input common mode range, high output swing under load and low power consumption. The device displays exceptional temperature stability and will operate over a wide range of supply voltages with little degradation of performance. The amplifier is intended for use in DC servo systems, high impedance analog computers, in low-level instrumentation applications and for the generation of special linear and nonlinear transfer functions. For full temperature range operation (-55°C to +125°C) see μA709 or μA709A data sheet.

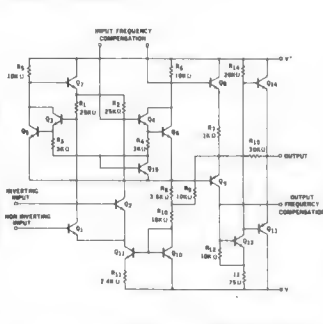
**ABSOLUTE MAXIMUM RATINGS**

Supply Voltage	±18 V
Internal Power Dissipation (Note 1)	250 mW
Differential Input Voltage	±5.0 V
Input Voltage	±10 V
Output Short-Circuit Duration (T <sub>A</sub> = 25°C)	5 sec
Storage Temperature Range	-65°C to +150°C
Operating Temperature Range	0°C to +70°C
Lead Temperature (Soldering, 60 sec)	300°C

**PHYSICAL DIMENSIONS**

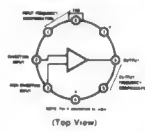


**SCHEMATIC DIAGRAM**

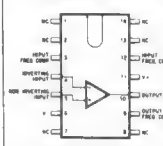


NOTE 1. Rating applies for ambient temperatures to +70°C

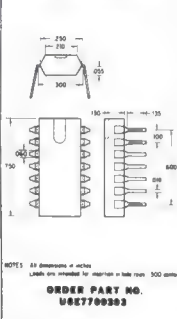
**CONNECTION DIAGRAM**



**CONNECTION DIAGRAM**



**PHYSICAL DIMENSIONS**



NOTE: All dimensions in inches.  
LEADS ARE INTENDED FOR MOUNTING IN SOLE TYPE 300 SERIES.

ORDER PART NO.  
U58770939X

111 FAIRCHILD DRIVE MOUNTAIN VIEW, CALIFORNIA 4151 962 5011 TWX 910 379 6435

**FAIRCHILD**  
SEMICONDUCTOR

COPYRIGHT FAIRCHILD SEMICONDUCTOR, 1967 • PRINTED IN U.S.A. 3435 280 87 10M  
MANUFACTURED UNDER ONE OR MORE OF THE FOLLOWING U.S. PATENTS 291877, 3015048, 3025589, 3064167, 3106559, 3117250, OTHER PATENTS PENDING

**Fig. 13.30** Example of data sheets for an op amp provided by a manufacturer. (Courtesy Fairchild Semiconductor.)

low offsets, high input impedance, etc. The amplifier is intended for such applications as dc servos, analog computers, low-level instrumentation, and the generation of special functions.

**2. Absolute maximum ratings** Included under this heading are such items as supply voltage, internal power dissipation, and operating temperature range. For reliable operation and long life, the absolute maximum ratings must never be exceeded.

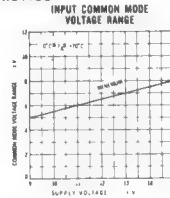
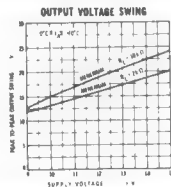
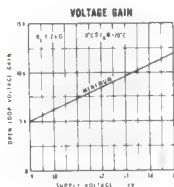
**3. Physical dimensions** Provides detailed dimensions of the package for mounting. Additional examples of IC packages are illustrated in Fig. 13.31.

**4. Schematic diagram** The schematic diagram is a circuit description of the op amp. The μA709C has some interesting features. The first differential stage (Q<sub>1</sub>, Q<sub>2</sub>)

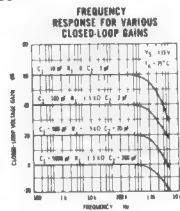
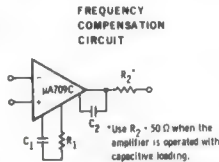
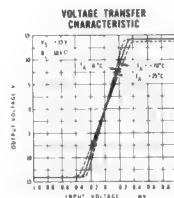
FAIRCHILD LINEAR INTEGRATED CIRCUITS  $\mu A709C$ ELECTRICAL CHARACTERISTICS ( $V_S = \pm 15V$ ,  $T_A = 25^\circ C$  unless otherwise specified)

Parameter	Conditions	Min.	Typ.	Max.	Units
Input Offset Voltage	$R_S \leq 10\text{ k}\Omega$ , $9V < V_S \leq \pm 15V$	2.0	7.5		mV
Input Offset Current		100	500		nA
Input Bias Current		0.3	1.5		$\mu A$
Input Resistance		50	250		k $\Omega$
Output Resistance			150		$\Omega$
Large-Signal Voltage Gain	$R_L = 2\text{ k}\Omega$ , $V_{out} = \pm 10V$	15,000	45,000		
Output Voltage Swing	$R_L = 10\text{ k}\Omega$	$\pm 12$	$\pm 14$		V
	$R_L = 2\text{ k}\Omega$	$\pm 10$	$\pm 13$		V
Input Voltage Range		$\pm 8.0$	$\pm 10$		V
Common Mode Rejection Ratio	$R_S \leq 10\text{ k}\Omega$	65	90		dB
Supply Voltage Rejection Ratio	$R_S \leq 10\text{ k}\Omega$	25	200		$\mu V/V$
Power Consumption		80	200		mW
Transient Response	$V_{in} = 20\text{ mV}$ , $R_L = 2\text{ k}\Omega$ , $C_L = 5000\text{ pF}$ , $R_1 = 1.5\text{ k}\Omega$ , $C_2 = 200\text{ pF}$ , $R_2 = 50\text{ }\Omega$		0.3		$\mu s$
Risetime					
Overshoot	$C_L \leq 100\text{ pF}$		10		%
The following specifications apply for $0^\circ C \leq T_A \leq +70^\circ C$					
Input Offset Voltage	$R_S \leq 10\text{ k}\Omega$ , $9V < V_S \leq \pm 15V$			10	mV
Input Offset Current				750	nA
Input Bias Current				2.0	$\mu A$
Large-Signal Voltage Gain	$R_L = 2\text{ k}\Omega$ , $V_{out} = \pm 10V$	12,000			
Input Resistance		55			k $\Omega$

## GUARANTEED ELECTRICAL CHARACTERISTICS



## TYPICAL PERFORMANCE CURVES



Fairchild cannot assume responsibility for use of any circuitry described other than circuitry entirely embodied in a Fairchild product. No other circuit patent licenses are implied.

Fig. 13.30 Continued.

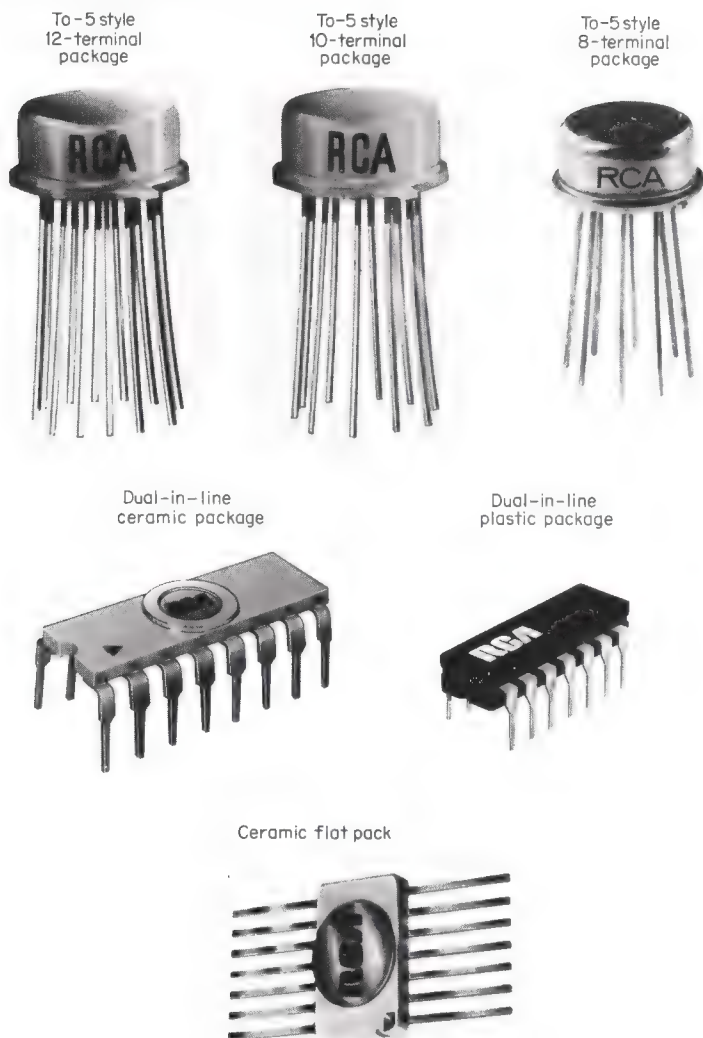
is cascaded to a second differential stage which uses *Darlington pairs* ( $Q_3$ ,  $Q_5$  and  $Q_4$ ,  $Q_8$ ) for the transistors. A Darlington pair provides a very high input impedance. (In some op amps, field-effect transistors are used in the first differential amplifier stage to achieve a high input impedance.) The output stage ( $Q_{13}$ ,  $Q_{14}$ ) is a complementary pair circuit.

**5. Connection diagram** Identifies the terminals on the package.

**6. Electrical characteristics** Provides ranges of parameter values such as input offset voltage, gain, and power consumption. Test conditions for the parameters must always be explicit. The *supply-voltage rejection ratio* is the ratio of a change in offset voltage for a corresponding change in a supply voltage.

**7. Guaranteed electrical characteristics** Curves show how voltage gain, output voltage swing, and input common-mode voltage range vary with the supply voltage.

### 13-18 Operational Amplifiers



**Fig. 13.31** Examples of commonly used IC packages. (Courtesy RCA.)

The *common-mode voltage range* is the maximum input voltage that may be applied to either the inverting or noninverting terminal of an op amp for safe operation.

**8. Typical performance curves** The *voltage-transfer characteristic* relates the output voltage to the input voltage over the operating temperature range of the op amp. The frequency compensation circuit and response curves are extremely important. As mentioned earlier, frequency compensation is required to ensure stable operation of the amplifier. It is noted that the closed-loop gain is dependent on the values of the capacitors and resistors used in the compensation network.

# Chapter 14

## Digital Circuit Fundamentals

### 14.1 INTRODUCTION

The development of the junction diode and transistor has been the chief catalyst in the rapid growth of computers and digital systems in general. Semiconductor devices can be made to act like nearly perfect switches, making them ideal for digital circuits. By the use of a few elementary circuits over and over again, all the necessary functions in the operation of a digital system are realized.

Because a switch is either ON or OFF, it is said to exhibit two stable states. (A switch, for this reason, is sometimes referred to as a *bistable device*.) For semiconductor switches to perform arithmetic and logic operations, it therefore becomes necessary to represent numbers by two digits, or *bits*, 0 and 1. In this manner, a 0 may be represented by a closed switch and a 1 by an open switch. Such a number system using only 0s and 1s, called the *binary number system*, is employed in digital systems.

In this chapter we consider first how numbers are expressed in the binary number system. It is also necessary to know how to convert a decimal number into a binary number and back again to decimal. Basic logic functions, such as AND, OR, NOT, NAND, and NOR, are examined next. These functions are described completely by the *truth table*, which is discussed. Operation of multivibrators is also considered. The chapter concludes with a study of the response of switching devices to rapidly changing waveforms, referred to as *dynamic response*.

### 14.2 BINARY NUMBERS

Because we are born with ten fingers, it is by sheer accident that we count by 10s. The *base*, or *radix*, of our number system is ten, referred to as the *decimal number system*. A zero and nine different symbols are needed to represent decimal numbers: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. In some past societies people used both their fingers and toes to count. It is not surprising to learn that the base of their number system was twenty. If we were born with only two hands and no fingers, it is a safe bet that we would count by 2s.

Number systems are *positional*. The value of a number depends not only on its symbols but also on their position in the number. For example, 367 and 673 are two numbers having the same symbols but different positions. One may regard a decimal number, such as 367, as a shorthand for

$$3 \times 10^2 + 6 \times 10^1 + 7 \times 10^0 = 300 + 60 + 7 = 367$$

where  $10^0 = 1$ ,  $10^1 = 10$ , and  $10^2 = 100$ . Thus, 367 is equal to 7 units ( $10^0$ ), 6 tens ( $10^1$ ), and 3 hundreds ( $10^2$ ). All decimal numbers are expressed in this manner, regardless of their magnitude.

## 14-2 Digital Circuit Fundamentals

The binary number system is also positional. For example, 1101 is equal to

$$1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 8 + 4 + 0 + 1 = 13$$

where  $2^0 = 1$ ,  $2^1 = 2$ ,  $2^2 = 4$ , and  $2^3 = 8$ . Thus, binary number 1101 is equal to 1 unit ( $2^0$ ), 0 twos ( $2^1$ ), 1 four ( $2^2$ ), and 1 eight ( $2^3$ ).

In comparing numbers having different base systems, it is customary to place a subscript, equal to the number base, after the number. Hence, one would write

$$13_{10} = 1101_2$$

**DECIMAL-TO-BINARY CONVERSION** To convert a whole decimal number to a binary number, the following procedure is used:

1. Divide the number by 2; the remainder is either a 0 or a 1.
2. Place the remainder to the right of the partial quotient obtained in step 1.
3. Divide the partial quotient of step 1 by 2, placing the remainder to the right of the new partial quotient.
4. Repeat the preceding steps until a quotient of zero is obtained.
5. The binary number is equal to the remainders arranged so that the first remainder is the least significant bit (LSB) and the last remainder is the most significant bit (MSB) of the binary number.

**example 14.1** Convert the following decimal numbers to binary numbers: (a) 13; (b) 103.

**solution** (a)

2)13	
2)6	1
2)3	0
2)1	1
0	1

} Remainders

Rearranging remainders,  $13_{10} = 1101_2$ . (This equality was verified in the preceding section.)

(b)

2)103	
2)51	1
2)25	1
2)12	1
2)6	0
2)3	0
2)1	1
0	1

Hence,  $103_{10} = 1100111_2$ . To verify the answer, the binary number is expanded into its constituent parts:  $1100111_2 = 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 1 \times 64 + 1 \times 32 + 0 \times 16 + 0 \times 8 + 1 \times 4 + 1 \times 2 + 1 \times 1 = 64 + 32 + 4 + 2 + 1 = 103_{10}$ .

**DECIMAL-TO-BINARY FRACTION CONVERSION** To convert a decimal fraction to a binary fraction, the following procedure is used:

1. Multiply the decimal fraction by 2.
2. If, as a result of step 1, a number equal to or greater than 1 is obtained, the 1 is placed to the right of the partial product. For example, multiplying 0.6 by 2 yields 1.2, which is a number greater than 1. The 1 is placed to the right of the partial product 0.2. If the result obtained is less than one, a zero is placed to the right of the partial product.
3. The partial product obtained in step 2 is multiplied by 2. The process is repeated until the partial product is 0 or the resulting binary fraction is to the required places of the binary point (which corresponds to the decimal point of a decimal fraction).
4. The 1s and 0s in the order obtained are equal to the binary fraction.

**example 14.2** Convert the following decimal fractions to binary fractions: (a) 0.375; (b) 0.68.

**solution** (a)

$2 \times 0.375 = 0.750$	0
$2 \times 0.750 = 1.500$	1
$2 \times 0.500 = 1.000$	1
$2 \times 0.000 = 0.000$	

Hence,  $0.375_{10} = 0.011_2$ . To verify the answer, 0.011 is expanded into its constituent parts:  $0.011 = 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} = 0/2 + 1/4 + 1/8 = 0.25 + 0.125 = 0.375$ .

(b)	$2 \times 0.68 = 1.36$	1
	$2 \times 0.36 = 0.72$	0
	$2 \times 0.72 = 1.44$	1
	$2 \times 0.44 = 0.88$	0
	$2 \times 0.88 = 1.76$	1

and so on. In this example the binary fraction is unending. To five *binary places*,  $0.68_{10} = 0.10101_2$ .

**BINARY-TO-DECIMAL CONVERSION** To convert a binary whole number to a decimal number, the following procedure is used:

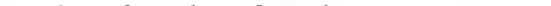
1. Multiply the most significant bit (MSB) by 2.
2. If the next to the MSB is a one, add a 1 to the partial product obtained in step 1; if a zero, add a 0.
3. Multiply the result obtained in step 2 by 2. Continue the process until the least significant bit (LSB) is included in the conversion.

**example 14.3** Convert the following binary numbers to decimal numbers: (a) 11010; (b) 110101.

<b>solution</b>	(a)	MSB	LSB
1	0000	0	0
2	0001	0	1
3	0010	0	0
4	0011	0	1
5	0100	1	0
6	0101	1	1
7	0110	1	0
8	0111	1	1
9	1000	1	0
10	1001	1	1
11	1010	0	0
12	1011	0	1
13	1100	0	0
14	1101	0	1
15	1110	1	0
16	1111	1	1

$$\begin{array}{ccccccc} 1 & & 1 & & 0 & & 1 & & 0 \\ \curvearrowright & & \curvearrowright & & \curvearrowright & & \curvearrowright & & \curvearrowright \\ 2 \times 1 + 1 = 3 & & 2 \times 3 + 0 = 6 & & 2 \times 6 + 1 = 13 & & 2 \times 13 + 0 = 26 \end{array}$$

Therefore,  $11010_2 = 26_{10}$ .

(b) 

Hence,  $110101_2 = 53_{10}$ .

**BINARY-TO-DECIMAL FRACTION CONVERSION** To convert a binary fraction to a decimal fraction, the following procedure is used:

1. Divide the LSB by 2.
2. If the next to the LSB is a 1, add it to the result of step 1.
3. Divide the result of step 2 by 2. If the next bit is a 1, add it to the result.
4. Continue the process until the binary point is reached.

**example 14.4** Convert the following binary fractions to decimal fractions: (a) 0.0111; (b) 0.01001.

**solution** (a) Dividing the LSB of 0.0111 by 2 yields  $\frac{1}{2} = 0.5$ . Because the next bit is a 1, it is added to 0.5. Hence, the first ratio,  $R_1$ , is

$$R_1 = 0.5 + 1 = 1.5$$

Dividing 1.5 by 2 and adding 1 yields

$$R_2 = \frac{1.5}{2} + 1 = 1.75$$

Dividing 1.75 by 2 and, because the next bit is a zero, we have

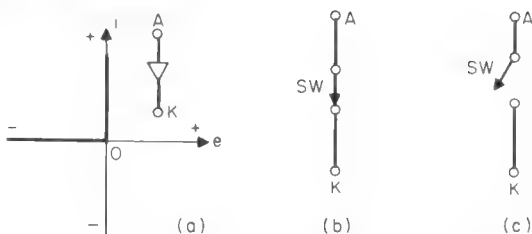
$$R_3 = \frac{1.75}{2} + 0 = 0.875$$

Finally, the binary point is reached and  $0.875/2 = 0.4375$ . Hence,  $0.0111_2 = 0.4375_{10}$ .

$$\begin{aligned}(b) \quad R_1 &= 0.5 + 0 = 0.5 \\ R_2 &= \frac{0.5}{2} + 0 = 0.25 \\ R_3 &= \frac{0.25}{2} + 1 = 1.125 \\ R_4 &= \frac{1.125}{2} + 0 = 0.5625 \\ R_5 &= \frac{0.5625}{2} = 0.28125\end{aligned}$$

Hence,  $0.01001_2 = 0.28125_{10}$ .



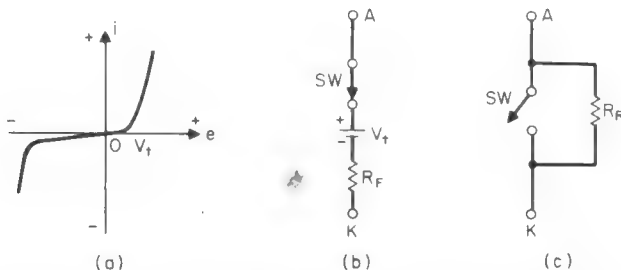


**Fig. 14.1** Ideal diode: (a) Current-voltage characteristics. (b) Closed-switch model for a forward-biased ideal diode. (c) Open-switch model for a reverse-biased ideal diode.

### 14.3 DIODE AS A SWITCH

As explained in Chap. 8, a junction diode when forward-biased conducts current and when reverse-biased it does not conduct current. *It behaves like a switch.*

Consider the characteristics of an ideal diode, shown in Fig. 14.1a. When forward-biased ( $A$  positive with respect to  $K$ ), the voltage across the ideal diode is zero, regardless of the current flowing. The forward-biased ideal diode, therefore, may be represented by the closed-switch model of Fig. 14.1b. On the other hand, when reverse-biased, the current is zero regardless of the magnitude of the reverse voltage. A reverse-biased ideal diode, therefore, can be represented by the open-switch model of Fig. 14.1c.



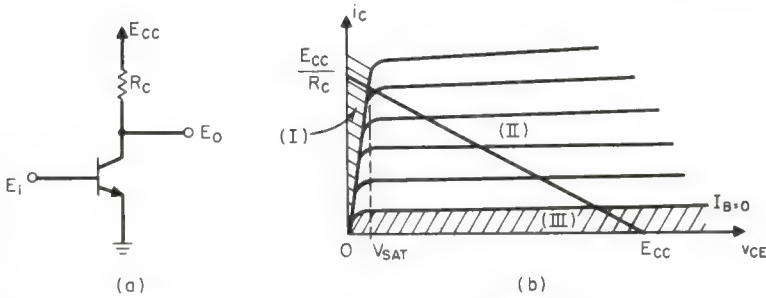
**Fig. 14.2** Physical junction diode: (a) Current-voltage characteristics. (b) Model for a forward-biased physical diode. (c) Model for a reverse-biased physical diode.

The characteristics of a physical junction diode are illustrated in Fig. 14.2a. When forward-biased, before appreciable current flows, the forward voltage must exceed the threshold voltage  $V_t$  (0.6 V for silicon at room temperature). In conduction, the forward resistance of the diode, which can be as low as a few ohms, is designated by  $R_F$ . A forward-biased physical diode, therefore, acts like a closed switch in series with a battery of  $V_t$  volts and a resistance of  $R_F$  ohms, as illustrated in Fig. 14.2b.

When reverse-biased (the reverse voltage must not exceed the breakdown voltage of the diode), the reverse resistance is extremely high, in the order of megohms. Designating the reverse resistance by  $R_R$ , the reverse-biased diode is represented by a switch in parallel with  $R_R$  shown in Fig. 14.2c. In practice, the models for the ideal diode are used often to represent a physical diode.

### 14.4 TRANSISTOR AS A SWITCH

Figure 14.3a is an elementary circuit of an npn junction transistor operating as a switch. (Field-effect transistor switches are considered in Chap. 15.) The configuration is that



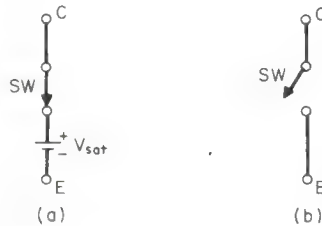
**Fig. 14.3** Transistor operating as a switch: (a) Elementary circuit employing a npn transistor in the common-emitter configuration. (If a pnp transistor is used, the polarity of  $E_{CC}$  is reversed.) (b) Load line superimposed on the collector characteristics. Regions I, II, and III are the saturation, active, and cutoff regions, respectively.

of a common-emitter amplifier. Input  $E_i$  is applied to the base of the transistor. The collector is connected to resistance  $R_C$  in series with the collector supply voltage  $E_{CC}$ . Output voltage  $E_o$  is taken across the collector and ground.

The load line superimposed on the collector characteristics is shown in Fig. 14.3b. One may define three regions of operation, indicated by Roman numerals I, II, and III.

**I. Saturation region** In the saturation region, both the collector-base and base-emitter junctions are *forward-biased*. The voltage across the collector and emitter is referred to as the *saturation voltage*,  $V_{sat}$ . Typically,  $V_{sat}$  for most switching transistors is 0.1 V or less.

When a transistor is in the saturation region, it is considered to be ON. It may be represented by the model of a closed switch in series with a battery of  $V_{sat}$  volts, as in Fig. 14.4a. A transistor that when ON is in saturation is referred to as a *saturated*



**Fig. 14.4** Models for a common-emitter transistor switch: (a) For the ON state. (b) For the OFF state.

*switch*. If a transistor is ON and not in the saturated region, it is referred to as a *non-saturated switch*. Because of greater efficiency, transistors are generally operated as saturated switches.

**II. Active region** In the active region, the collector-base junction is *reverse-biased* and the base-emitter junction is *forward-biased*. The Q point of an amplifier, for example, is located in the active region of the collector characteristics.

**III. Cutoff region** In the cutoff region, both the collector-base and base-emitter junctions are *reverse-biased*. When in the cutoff region, the transistor is considered to be OFF. It may be represented by the model of an open switch of Fig. 14.4b.

A transistor operated as a saturated switch is either in the saturation (ON) or cutoff (OFF) regions. In switching a transistor from ON to OFF (or from OFF to ON), it makes a transition through the active region. A *fast* switching transistor makes the transition rapidly.

14.5 BASIC LOGIC CIRCUITS

There are three basic logic circuits which, when used in various combinations, perform the arithmetic and control functions in a digital computer or system. These circuits are the OR gate, AND gate, and the INVERTER (NOT).

**OR GATE** The logic symbol for a two-input OR gate is given in Fig. 14.5. Output  $X$  is obtained when  $A$  or  $B$ , or both  $A$  and  $B$ , are present at the input. The preceding statement can be expressed by a logic equation as

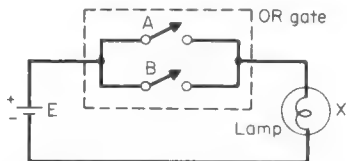
$$X = A + B \tag{14.1}$$

where the plus sign (+) is interpreted as OR. Because the binary number system is used in logic systems, inputs  $A$  and  $B$ , and output  $X$  can assume only the values of 0 and 1.

An elementary example of an OR gate is the two parallel manual switches,  $A$  and  $B$  of Fig. 14.6. The input to both switches is voltage source  $E$ ; the output  $X$  is a glowing lamp. The lamp glows if switch  $A$  or  $B$  or both  $A$  and  $B$  are closed.



**Fig. 14.5** Logic symbol for a two-input OR gate ( $X = A + B$ ). OR gates having more than two inputs are also possible.



**Fig. 14.6** An example of an OR gate consisting of two parallel manual switches. The lamp glows ( $X = 1$ ) if switch  $A$ , or  $B$ , or  $A$  and  $B$  are closed.

The operation of a logic circuit can be described by a *truth table*. In a truth table, all possible combinations of inputs and their corresponding outputs are displayed in tabular form. Inputs  $A$  and  $B$  can assume only four possible combinations of values: 00, 01, 10, and 11. These values and their corresponding output  $X$  are listed in the truth table for the two-input OR gate in Fig. 14.7. It is read as an output  $X$  exists when  $A = 0$  or  $B = 1$ ;  $A = 1$  or  $B = 0$ ; or  $A = B = 1$ . No output exists when  $A = B = 0$ .

The number of different possible combinations assumed by  $n$  inputs is expressed by

$$\text{Combinations} = 2^n \tag{14.2}$$

As we saw in the preceding paragraph, for two inputs  $A$  and  $B$  four combinations were possible ( $2^n = 2^2 = 4$ ).

**example 14.5** Draw the truth table for a three-input ( $A, B, C$ ) OR gate.

A	B	X
0	0	0
0	1	1
1	0	1
1	1	1

**Fig. 14.7** Truth table for a two-input OR gate.

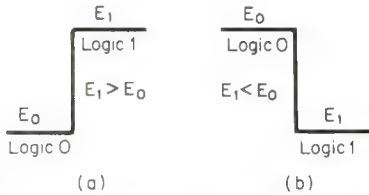
A	B	C	X
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

**Fig. 14.8** Truth table for a three-input OR gate. (See Example 14.5.)

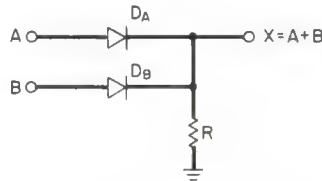
**solution** Because  $n = 3$ , by Eq. (14.2),  $2^3 = 8$  combinations are possible. These are: 000, 001, 010, 011, 100, 101, 110, and 111. (A simple way of listing the combinations is to write the binary numbers corresponding to decimal numbers 0 to  $2^n - 1$ .) The combinations and the corresponding values of  $X$  for the three-input OR gate are listed in the truth table of Fig. 14.8.

**Positive and negative logic** In *positive logic*, the voltage that represents a 1 is greater than the voltage that represents a 0. The reverse is true in *negative logic*: A 1 is represented by a voltage less than the voltage denoting a 0. Letting  $E_1$  be the voltage for a 1 and  $E_0$  be the voltage for a 0, examples of positive and negative logic are illustrated in Fig. 14.9. The 1 and 0 are commonly referred to as *logic 1* and *logic 0*.

**OR gate circuit** An example of a two-input OR gate (for positive logic) using diodes is illustrated in Fig. 14.10. (In negative logic the diodes are reversed.) The circuit is an example of *diode logic*.



**Fig. 14.9** Example of positive and negative logic levels: (a) Positive logic:  $E_1 > E_0$ . (b) Negative logic:  $E_1 < E_0$ .



**Fig. 14.10** An example of a two-input OR gate using diodes for positive logic. (In negative logic, the diode connections are reversed.)

Let  $E_1 = 5 \text{ V}$  and  $E_0 = 0 \text{ V}$ . Assuming ideal diodes, if  $E_A = E_B = 0 \text{ V}$ , both diodes are nonconducting, and the output  $= 0 \text{ V}$  ( $X = 0$ ). If  $E_A = 5 \text{ V}$  and  $E_B = 0 \text{ V}$ , diode  $D_A$  conducts and diode  $D_B$  is reverse-biased; the output, therefore, equals  $5 \text{ V}$  ( $X = 1$ ). If  $E_B = 5 \text{ V}$  and  $E_A = 0 \text{ V}$ , diode  $D_B$  now conducts and  $D_A$  is reverse-biased; output equals  $5 \text{ V}$ . For  $E_A = E_B = 5 \text{ V}$ , both diodes conduct and the output is at  $5 \text{ V}$  ( $X = 1$ ).

Because of noise in digital systems, it is possible that voltages  $E_A$  and  $E_B$  are unequal. Suppose that  $E_A = 6 \text{ V}$  and  $E_B = 5 \text{ V}$  for a logic 1. If  $E_A = 6 \text{ V}$  and  $E_B = 0 \text{ V}$ , the output equals  $6 \text{ V}$ . If  $E_A = 5 \text{ V}$  and  $E_B = 0 \text{ V}$ , the output will be at  $5 \text{ V}$ . If both inputs are applied simultaneously,  $E_A = 6 \text{ V}$  and  $E_B = 5 \text{ V}$ , the output is equal to the *larger of the inputs*.

Assume that indeed  $E_A = 6 \text{ V}$  and  $E_B = 5 \text{ V}$ . The output is  $6 \text{ V}$ . Because the anode of diode  $D_B$  is at  $5 \text{ V}$  and the cathode at  $6 \text{ V}$ , diode  $D_B$  is reverse-biased and does not conduct.



**Fig. 14.11** Logic symbol for a two-input AND gate ( $X = AB$ ). AND gates having more than two inputs are also possible.

**AND GATE** The logic symbol for a two-input AND gate is shown in Fig. 14.11. Output  $X = 1$  when inputs  $A$  and  $B$  are present; otherwise,  $X = 0$ . Expressed by a logic equation,

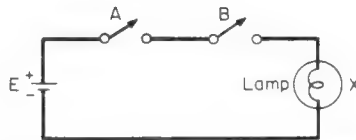
$$X = A \times B \quad (14.3a)$$

where the  $\times$  is interpreted as AND. As in algebra, the  $\times$  may be eliminated and Eq. (14.3a) expressed by

$$X = AB \quad (14.3b)$$

A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

**Fig. 14.12**  
Truth table  
for a two-  
input AND  
gate.

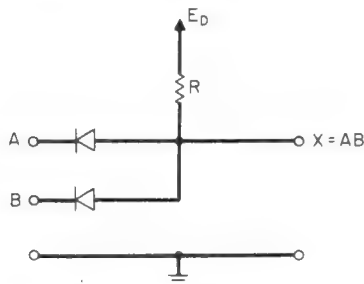


**Fig. 14.13** An example of an AND operation. The lamp glows ( $X = 1$ ) only if switches A and B are closed.

The truth table for a two-input AND gate is given in Fig. 14.12. It is seen that an output exists ( $X = 1$ ) when *both* inputs equal 1; otherwise,  $X = 0$ . For a three-input AND gate,  $X = 1$  only when all three inputs equal logic 1, and so on.

An elementary example of an AND operation is illustrated in Fig. 14.13. Switches A and B are connected in series with source E and the lamp. The lamp glows ( $X = 1$ ) when switches A and B are closed. If either or both switches are open, the lamp does not glow ( $X = 0$ ).

**AND gate circuit** An example of a two-input AND gate for positive logic using diodes is shown in Fig. 14.14. Resistance R is returned to voltage  $E_D$  which is equal to or greater than  $E_1$ . (For negative logic, the diodes and polarity of  $E_D$  are reversed.)



**Fig. 14.14** An example of a two-input AND gate for positive logic. (In negative logic, the diode connections and supply voltage  $E_D$  are reversed.)

Assume ideal diodes,  $E_1 = 5$  V,  $E_0 = 0$  V, and  $E_D = 5$  V. If either  $E_A$  or  $E_B = 0$  V, diode  $D_A$  or  $D_B$  conducts; the output is zero volts ( $X = 0$ ). If  $E_A = E_B = 5$  V, both diodes are reverse-biased and the output equals  $E_D = 5$  V ( $X = 1$ ).

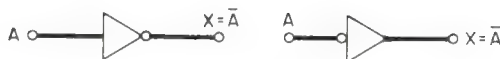
When  $E_A$  and  $E_B$  for logic 1 are unequal, the output is equal to the *lesser* voltage. Assume that  $E_A = 5$  V and  $E_B = 3$  V. Diode  $D_B$  becomes forward-biased because its anode was initially at  $E_D = 5$  V and its cathode at  $E_B = 3$  V. The output is equal to 3 V ( $X = 1$ ). Since the anode of diode  $D_A$  is now at 3 V and its cathode at  $E_A = 5$  V, diode  $D_A$  is reverse-biased.

**INVERTER** The logic symbol for the INVERTER, also referred to as a NOT, is given by either of the symbols of Fig. 14.15. Output X is equal to the *not* of A. Expressed by a logic equation,

$$X = \overline{A} \quad (14.4)$$

where the bar over A indicates a NOT operation. If  $A = 1$ , output  $X = 0$ ; if  $A = 0$ ,  $X = 1$ . The truth table for the inverter is given in Fig. 14.16.

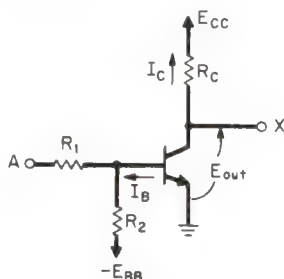
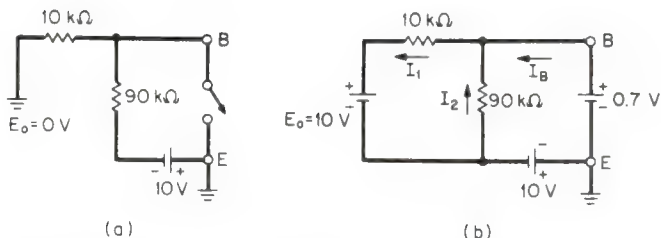
**INVERTER circuit** An example of an inverter using an npn junction transistor is illustrated in Fig. 14.17. Resistance  $R_1$  is in series with the transistor base and input A.

Fig. 14.15 Logic symbols for the INVERTER ( $X = \bar{A}$ ).

A	X
0	1
1	0

Fig. 14.16 Truth table for an INVERTER.

Resistance  $R_2$  is connected between the base and the negative-voltage supply,  $-E_{BB}$  volts. This ensures that for  $A = E_0$  (logic 0), the base-emitter junction is reverse-biased and the output  $= E_{CC}$  volts (logic 1). When  $A = E_1$  (logic 1), transistor  $Q$  conducts and the output is equal to  $V_{sat}$  volts (logic 0).

Fig. 14.17 An example of an INVERTER circuit using a npn transistor. For a pnp transistor the polarities of  $E_{CC}$  and  $E_{BB}$  are reversed.Fig. 14.18 Models for the input (base) circuit of a transistor switch: (a) For  $E_0 = 0$  V. (b) for  $E_1 = 10$  V. (See Example 14.6.)

**example 14.6** For the INVERTER of Fig. 14.17,  $E_{CC} = 10$  V,  $-E_{BB} = -10$  V,  $R_C = 1$  kΩ,  $R_1 = 10$  kΩ, and  $R_2 = 90$  kΩ. Logic signals  $E_0 = 0$  V and  $E_1 = 10$  V. (a) Verify that the circuit behaves as an INVERTER. (b) Determine the minimum dc current gain for the transistor,  $h_{FE, min}$  to ensure that the transistor operates as a saturated switch. Assume that  $V_{sat} = 0.1$  V and that the base-emitter voltage is 0.7 V when the transistor is ON.

**solution** (a) For  $E_0 = 0$  V, the base-emitter junction is reverse-biased and is represented by the model of an open switch, shown in Fig. 14.18a. The voltage across the base-emitter junction, by voltage division, is  $-10 \times 10 / (10 + 90) = -1$  V. A reverse bias of  $-1$  V is more than necessary for a transistor to be OFF (nonconducting).

For  $E_1 = 10$  V, the model for the base circuit is illustrated in Fig. 14.18b. The base-emitter junction is represented by a 0.7-V battery. To calculate the base current  $I_B$ , superposition is used. Setting  $E_{BB}$  to zero, the current owing to  $E_1$  is  $I_1 = (10 - 0.7) / 10 = 0.93$  mA. Setting  $E_1$  to zero, the current due to  $E_{BB}$  is  $I_2 = (10 + 0.7) / 90 = 0.12$  mA. The base current is equal to the difference of  $I_1$  and  $I_2$ :  $I_B = I_1 - I_2$ . Hence,  $I_B = 0.93 - 0.12 = 0.81$  mA.



14-10 Digital Circuit Fundamentals

From Fig. 14.17, collector current  $I_C$  is equal to the difference of  $E_{CC}$  and  $V_{sat}$  divided by  $R_C$ :  $I_C = (E_{CC} - V_{sat})/R_C$ . Substitution of the given values in the equation yields  $I_C = (10 - 0.1)/1 = 9.9$  mA. When input  $A = E_1$  (10 V), the output equals  $V_{sat}$  (logic 0) if  $I_B = 0.81$  mA.

(b) The minimum dc current gain is expressed by  $h_{FE,min} = I_C/I_B$ . From part (a),  $h_{FE,min} = 9.9/0.81 = 12.2$ . Transistor  $Q$ , therefore, must have a dc current gain of at least 12.2 to ensure that it operates as a saturated switch.

14.6 HALF-ADDER

To demonstrate the usefulness of the AND, OR, and INVERTER circuits, these building blocks are used to construct a basic component of a binary adder, the half-adder (HA). Besides addition, the binary adder can be used for subtraction, multiplication, and division. In subtraction, the subtrahend is inverted and added to the minuend. Multiplication may be performed by repeated addition, and division by repeated subtraction. The binary adder is indeed a universal circuit.

The addition table for binary numbers is extremely simple, as illustrated in Fig. 14.19. We see that  $0 + 0 = 0$ ;  $0 + 1 = 1$ ;  $1 + 0 = 1$ ; and  $1 + 1 = 10$  (decimal 2). The heart of the binary adder is the half-adder, shown symbolically in Fig. 14.20. It has two inputs  $A$  and  $B$ , and two outputs  $C$  (carry) and  $S$  (sum).

		Augend			
		+			
		0	1		
Addend	0	0	1	Sum	
	1	0	10		

Fig. 14.19 Addition table for binary numbers.



Fig. 14.20 Symbol for a half-adder. Bits to be added are  $A$  and  $B$ . Output  $S$  is the sum bit and  $C$  is the carry bit.

A	B	C	S
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Fig. 14.21 Truth table for a half-adder.

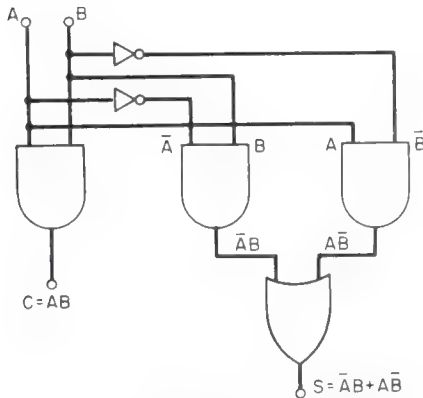


Fig. 14.22 A half-adder realized with AND and OR gates and INVERTERS.

The truth table for the half-adder is given in Fig. 14.21. A sum bit results when  $A = 0$  and  $B = 1$ , or  $A = 1$  and  $B = 0$ . When  $A = 1$  and  $B = 1$ , a 0 results in the sum column and a 1 in the carry column.

The preceding results may be summarized by two logic equations. Sum  $S = 1$  when  $A = 0$  and  $B = 1$  or  $A = 1$  and  $B = 0$ . Carry  $C = 1$  when  $A$  and  $B = 1$ . Hence,

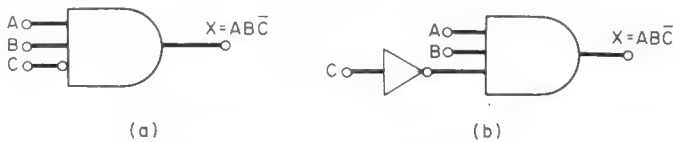
$$S = \bar{A}B + A\bar{B} \quad (14.5a)$$

$$C = AB \quad (14.5b)$$

Using AND and OR gates, and INVERTERS, the half-adder may be realized as shown in Fig. 14.22.

## 14.7 INHIBIT AND EXCLUSIVE OR GATES

The logic symbol for a three-input INHIBIT gate is illustrated in Fig. 14.23a. It is basically an AND gate with one input ( $C$ ) negated by an INVERTER, as shown in Fig. 14.23b. This is indicated on the symbol by a circle drawn at input  $C$ .



**Fig. 14.23** An INHIBIT gate: (a) Logic symbol ( $X = ABC$ ). (b) It consists of an AND gate with one input negated by an INVERTER. INHIBIT gates having more than three inputs, as well as only two inputs, are possible.

The operation of an INHIBIT gate is such that  $X = 1$  when  $A = B = 1$  and  $C = 0$ ; if  $C$  also equals 1,  $X = 0$ . Input  $C$  therefore acts to *inhibit* the gate. If  $C = 0$ , it *enables* the gate, permitting it to perform an AND operation. The logic equation for the INHIBIT gate is

$$X = AB\bar{C} \quad (14.6)$$

The truth table is given in Fig. 14.24. Because three inputs are present,  $n = 3$ . By Eq. (14.2),  $2^3 = 8$  possible combinations of  $A$ ,  $B$ , and  $C$  exist. It is seen that  $X = 1$  only if  $A = 1$ ,  $B = 1$ , and  $C = 0$ .

The logic symbol for an EXCLUSIVE OR gate is shown in Fig. 14.25. For this gate,  $X = 1$  if  $A = 1$ ,  $B = 0$ , or  $A = 0$ ,  $B = 1$ . If  $A = B = 0$  or  $A = B = 1$ , the output is zero. Expressed by a logic equation,

$$X = \bar{A}B + A\bar{B} \quad (14.7)$$

A	B	C	X
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	0

**Fig. 14.24** Truth table for a three-input INHIBIT gate.



**Fig. 14.25** Logic symbol for an EXCLUSIVE OR gate ( $X = \bar{A}B + A\bar{B}$ ).

Equation (14.7) is identical with Eq. (14.5a) for the sum  $S$  of a half-adder. The circuit for  $S$  in Fig. 14.22 can therefore be used for realizing an EXCLUSIVE OR gate.

The truth table is illustrated in Fig. 14.26. Comparing this table with the truth table for the OR gate of Fig. 14.7, it is seen that in the OR gate,  $X = 1$  even if  $A = B = 1$ . The OR gate, therefore, is sometimes referred to as an INCLUSIVE OR, because it provides an output when all inputs are at a logic 1.

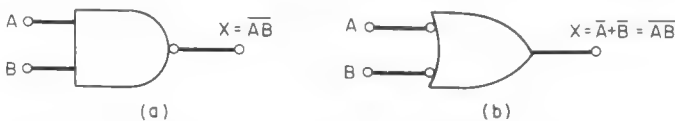
A	B	X
0	0	0
0	1	1
1	0	1
1	1	0

**Fig. 14.26**  
Truth table  
for an EX-  
CLUSIVE  
OR gate.

14.8 NAND AND NOR GATES

Two universal gates, the NAND and NOR, are covered in this section. With either of these gates, the OR, AND, and NOT functions can be realized.

**NAND GATE** Logic symbols for the NAND gate are illustrated in Fig. 14.27. The NAND gate may be regarded as an AND gate followed by an INVERTER, as shown in



**Fig. 14.27** Logic symbols for a two-input NAND gate ( $X = \overline{AB} = \overline{A} + \overline{B}$ ). NAND gates with more than two inputs are also possible.

Fig. 14.28. Output  $X = 1$  when  $A = B = 0$ ; otherwise  $X = 0$ . Expressed by a logic equation,

$$X = \overline{AB} \tag{14.8a}$$

which is read as “ $X$  is equal to the *not* of  $A$  and  $B$ .” By a form of a very useful theorem, De Morgan’s theorem, the *not* of  $A$  and  $B$  is equal to the *not* of  $A$  or the *not* of  $B$ :  $\overline{AB} = \overline{A} + \overline{B}$ . The logic equation for NAND operation may therefore also be expressed by

$$X = \overline{A} + \overline{B} \tag{14.8b}$$



**Fig. 14.28** The NAND gate viewed as an AND gate followed by an INVERTER.

A	B	X
0	0	1
0	1	1
1	0	1
1	1	0

**Fig. 14.29**  
Truth  
table for a  
two-input  
NAND  
gate.

Based on Eq. (14.8b), an alternative symbol for the NAND gate is given in Fig. 14.27b, which shows negated inputs to an OR gate.

The truth table for a two-input NAND gate for positive logic is provided in Fig. 14.29. It is seen that  $X = 0$  when  $A = B = 1$ ; for all other inputs,  $X = 1$ .

If both inputs  $A$  and  $B$  are connected together, the NAND gate functions as an INVERTER. An INVERTER following a NAND yields an AND gate, as illustrated in Fig. 14.30. A *not of a not* ( $\overline{\overline{AB}}$ ), indicated by two horizontal bars above  $AB$ , yields the function itself,  $AB$ . For negative logic, where a logic 1 is represented by a voltage that is less positive than for a logic 0, the positive NAND gate behaves as an OR gate.

**example 14.7** Verify that for negative logic the positive NAND gate behaves as an OR gate.

**solution** In the truth table of Fig. 14.29, a 0 becomes a 1 and a 1 becomes a 0 for inputs  $A$  and  $B$  in negative logic. The new truth table, therefore, appears as in Fig. 14.31. This is identical with the truth table for the OR gate of Fig. 14.7.



**Fig. 14.30** A two-input AND gate realized by a NAND gate followed by a second NAND gate acting as an INVERTER.

A	B	X
1	1	1
1	0	1
0	1	1
0	0	0

**Fig. 14.31** Truth table for a two-input positive NAND gate functioning as an OR gate for negative logic. (See Example 14.7.)

**NOR GATE** Logic symbols for the NOR gate are given in Fig. 14.32. The NOR gate may be viewed as an OR gate followed by an INVERTER, illustrated in Fig. 14.33. Output  $X = 1$  only when  $A = B = 0$ ; otherwise  $X = 0$ . Expressed by a logic equation,

$$X = \overline{A + B} \quad (14.9a)$$

which is read as “ $X$  is equal to the *not of A or B*.”

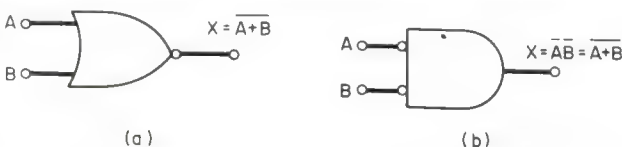
By another form of De Morgan’s theorem,  $\overline{A + B}$  is equal to the *not of A and the not of B*. Therefore  $X$  also equals  $\overline{A} \times \overline{B}$ :

$$X = \overline{A} \times \overline{B} \quad (14.9b)$$

Based on Eq. (14.9b), an alternative symbol for the NOR gate is given in Fig. 14.32b, which shows negated inputs to an AND gate.

The truth table for a two-input NOR gate for positive logic is illustrated in Fig. 14.34. It is seen that output  $X = 1$  only for  $A = B = 0$ ; for all other pairs of inputs,  $X = 0$ .

If both inputs are tied together, the NOR gate acts as an INVERTER. An INVERTER



**Fig. 14.32** Logic symbols for a two-input NOR gate ( $X = \overline{A + B} = \overline{A} \times \overline{B}$ ). NOR gates with more than two inputs are also possible.

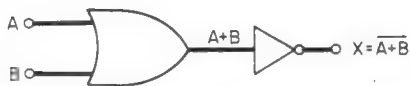


Fig. 14.33 A NOR gate viewed as an OR gate followed by an INVERTER.

A	B	X
0	0	1
0	1	0
0	1	0
1	1	0

Fig. 14.34 Truth table for a two-input NOR gate.

following a NOR yields an OR gate, as shown in Fig. 14.35. For negative logic, the positive NOR gate acts as an AND gate.

**example 14.8** Verify that for negative logic the positive NOR gate acts as an AND gate.

**solution** In the truth table of Fig. 14.34, a 0 becomes a 1 and a 1 becomes a 0 for inputs A and B in negative logic. The resulting truth table appears in Fig. 14.36. This is identical with the truth table for the AND gate of Fig. 14.12.



Fig. 14.35 A two-input OR gate realized by a NOR gate followed by a second NOR gate acting as an INVERTER.

A	B	X
1	1	1
1	0	0
0	1	0
0	0	0

Fig. 14.36 Truth table for a two-input positive NOR gate functioning as an AND gate for negative logic. (See Example 14.8.)

## 14.9 MULTIVIBRATORS

There are three kinds of multivibrators: the *bistable*, *monostable*, and *astable*. Their fundamental behavior and properties are explained initially with the aid of the block diagrams shown in Fig. 14.37.

a. The block diagram of a bistable multivibrator (MV), also referred to as a *flipflop*, *binary*, and an *Eccles-Jordan circuit*, is illustrated in Fig. 14.37a. The bistable MV has two inputs, labeled 1 and 2, and two outputs, X and  $\bar{X}$ . Output  $\bar{X}$  (not X) is said to be the *complement* of X. If X is a logic 1,  $\bar{X}$  is a logic 0; if X = 0, then  $\bar{X} = 1$ .

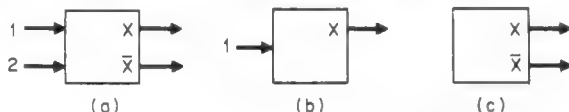


Fig. 14.37 Block diagrams of multivibrators: (a) Bistable. (b) Monostable. (c) Astable.

Assume that  $X = 1$ . The bistable MV remains in this state until a trigger is applied to, say, input 1; then  $X = 0$ . Output  $X = 0$  until a trigger is applied to terminal 2; then  $X = 1$  once again. Thus, the bistable MV exhibits two stable dc states:  $X = 1$  and  $X = 0$ . It remains in one of these two states until a trigger is applied to either terminal 1 or 2 to change its state.

The bistable MV exhibits *memory*. Its output depends not only on the present input, but also on which input terminal the previous trigger was applied. Such a circuit is referred to as *sequential*. The OR, AND, etc., gates depend only on their present inputs and are said to be *combinational*.

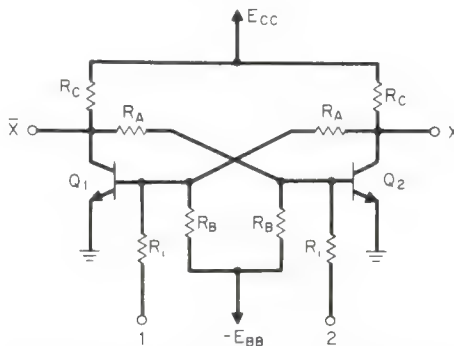
The bistable MV is used in *shift registers* and *counters*. These are two important components employed in digital computers and systems. Shift registers are used to store information, such as data or an instruction. The counter, as its name implies, counts pulses.

b. The block diagram of a monostable MV, commonly referred to as a *one-shot*, is illustrated in Fig. 14.37b. It has one input and one output terminal. With no trigger applied, the output voltage is approximately 0 V (logic 0). It remains in this stable dc state until a trigger is applied to the input terminal. As a result of the trigger, a well-defined rectangular pulse appears at  $X$ . At the termination of the rectangular pulse, the output returns to its stable state, logic 0.

In addition to generating rectangular pulses, the one-shot is also used as a delay. If, for example, the output pulse of a one-shot is applied to terminal C of the INHIBIT gate of Fig. 14.23a, the output of the gate is delayed for a time equal to the length of the rectangular pulse.

c. The block diagram of an astable, or *free-running*, MV, is given in Fig. 14.37c. There are two output terminals,  $X$  and  $\bar{X}$ , and no input terminals. At the outputs are rectangular-type waveforms. The astable MV is an example of a *nonsinusoidal*, or *relaxation*, oscillator. It is used as a *clock* in digital computers to ensure that operations, such as addition, are synchronized with other operations in the system.

**BISTABLE MV** The basic circuit of a *collector-coupled* bistable MV employing npn transistors is illustrated in Fig. 14.38. Note the symmetrical nature of the circuit. The



**Fig. 14.38** Basic circuit of a collector-coupled bistable MV employing npn transistors. (For pnp transistors, the polarities of  $E_{CC}$  and  $E_{BB}$  for this and other multivibrator circuits are reversed.)

collector of one transistor is coupled to the base of the other transistor by resistor  $R_A$ . Resistor  $R_B$  is connected to the base and supply source,  $-E_{BB}$  volts. Output  $X$  is taken at the collector of transistor  $Q_2$  and  $\bar{X}$  at the collector of transistor  $Q_1$ .

Assume that  $X = 1$  and, therefore,  $\bar{X} = 0$ . In this state,  $Q_1$  is conducting (ON) and  $Q_2$  is nonconducting (OFF). The base-emitter voltage of  $Q_1$  is approximately 0.7 V and the base-emitter voltage of  $Q_2$  is negative, ensuring that it is cut off.

To change the state of the flipflop, a negative trigger is applied to terminal 1. (For best triggering of multivibrators, the trigger is selected to *turn off* the ON transistor.)



Transistor  $Q_1$  is turned off and its collector voltage rises toward  $E_{CC}$  volts. The increasing collector voltage is coupled by resistor  $R_A$  to the base of  $Q_2$ . Its base-emitter voltage, therefore, becomes less negative and is finally equal to 0.7 V. Transistor  $Q_2$  is now ON ( $X = 0$ ), and  $Q_1$  is cut off ( $\bar{X} = 1$ ).

To obtain  $X = 1$  and  $\bar{X} = 0$ , a negative trigger is applied to terminal 2. The chain of events described in the preceding paragraph occurs, forcing  $Q_1$  to go ON and  $Q_2$  to go OFF. Now,  $X = 1$  and  $\bar{X} = 0$ .

**MONOSTABLE MV** The basic circuit of a collector-coupled monostable MV (one-shot) using npn transistors is shown in Fig. 14.39. As in the bistable MV, the collector of  $Q_2$  is coupled to the base of  $Q_1$  by resistor  $R_A$ . The collector of  $Q_1$ , however, is coupled to the base of  $Q_2$  by coupling capacitor  $C$ .

In its dc stable state,  $Q_2$  is ON ( $X = 0$ ). To ensure that indeed  $Q_2$  is ON, the maximum value of  $R$ ,  $R_{\max}$ , is equal to the net voltage across it ( $E_{CC} - \text{base-emitter voltage } E_{BE2}$ ) divided by base current  $I_{BS}$  required for  $Q_2$  to be in saturation:

$$R_{\max} = \frac{E_{CC} - E_{BE2}}{I_{BS}} \quad (14.10)$$

To aid us in understanding the operation of the one-shot, waveforms at the bases and collectors of  $Q_1$  and  $Q_2$  are shown in Fig. 14.40. For time less than  $t_1$ , the one-shot is in

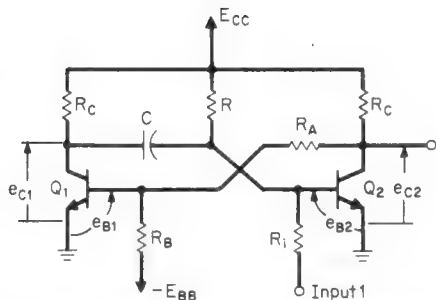


Fig. 14.39 Basic circuit of a collector-coupled monostable MV employing npn transistors.

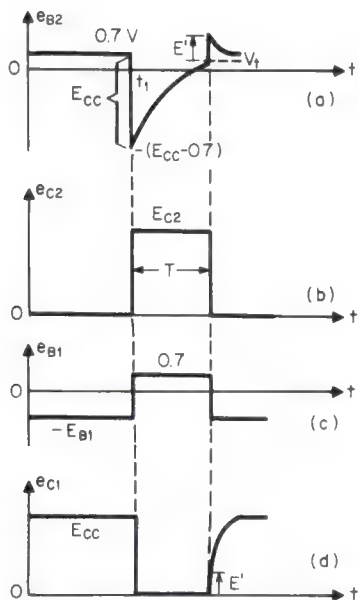


Fig. 14.40 Waveforms at the bases and collectors of transistors in a collector-coupled monostable MV: (a) Base of  $Q_2$ . (b) Collector of  $Q_2$ . (c) Base of  $Q_1$ . (d) Collector of  $Q_1$ .

its stable state where  $Q_1$  is OFF and  $Q_2$  is ON. The base-emitter voltage of  $Q_2$ ,  $e_{B2}$ , is at approximately 0.7 V. Assuming that  $V_{sat} = 0$ , the collector-emitter voltage of  $Q_2$  is, therefore,  $e_{C2} = 0$ . Because  $Q_1$  is OFF, its base-emitter voltage is negative ( $-E_{B1}$  volts) and its collector-emitter voltage,  $e_{C2}$ , is equal to the supply voltage,  $E_{CC}$  volts.

At  $t = t_1$  assume that a negative trigger is applied to the input terminal. Transistor  $Q_1$  turns ON and  $Q_2$  turns OFF. For  $V_{sat} = 0$ , the voltage at the collector of  $Q_1$  drops

by  $E_{CC}$  volts. Because the voltage across a capacitor cannot change instantaneously, the drop in  $E_{CC}$  volts is transmitted to the base of  $Q_2$ . Hence, at  $t = t_1$ ,  $e_{B2} = -(E_{CC} - 0.7)$  volts, and  $Q_2$  is cut off.

Voltage  $e_{B2}$  begins to rise exponentially toward  $E_{CC}$  volts. When it is equal to the threshold voltage  $V_i$ , transistor  $Q_2$  turns ON and  $Q_1$  turns OFF. At the base of  $Q_2$  an overshoot of  $E'$  volts occurs. Because of the coupling capacitor  $e_{C1}$  rises instantaneously by  $E'$  volts. During the time  $Q_2$  is OFF, its collector-emitter voltage,  $e_{C2}$ , is equal to  $E_{CC}$  volts which is slightly less than the supply voltage,  $E_{CC}$  volts.

The length, or width, of the rectangular pulse  $T$  generated at the collector of  $Q_2$  is given by

$$T \approx 0.69RC \quad (14.11)$$

where  $T$  is in seconds,  $R$  in ohms, and  $C$  in farads. Equation (14.11) is an excellent approximation for  $T$  if  $E_{CC}$  is much greater than 0.7 V. The length of the negative trigger that initiates the one-shot is generally a small fraction of  $T$ .

**ASTABLE MV** The basic circuit of a collector-coupled astable MV using npn transistors is shown in Fig. 14.41. Note that the collector of each transistor is coupled by capacitor  $C$  to the base of the other transistor. The astable MV, therefore, has no dc stable states.

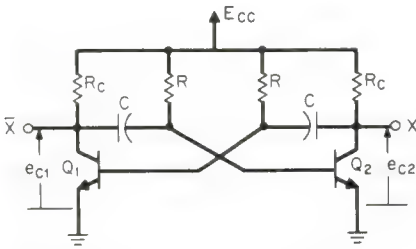


Fig. 14.41 Basic circuit of a collector-coupled astable MV employing npn transistors.

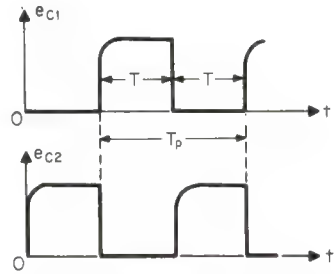


Fig. 14.42 Waveforms at the collectors of  $Q_1$  and  $Q_2$  of a collector-coupled astable MV.

Referring to the waveforms at the collectors of  $Q_1$  and  $Q_2$ ,  $e_{C1}$  and  $e_{C2}$ , respectively (Fig. 14.42), it is seen that each transistor is always generating a waveform. The shape of the waveforms is similar to that generated at the collector of  $Q_1$  in the one-shot. The length  $T$  indicated in Fig. 14.42 is equal to  $0.69RC$ , as for the one-shot. The total period  $T_P$  is therefore equal to  $2T$ :

$$T_P = 1.38RC \quad (14.12a)$$

The frequency  $f$  is one divided by the total period. Hence the frequency of the generated square wave, in hertz, is

$$f = \frac{1}{1.38RC} \quad (14.12b)$$

## 14.10 DYNAMIC RESPONSE OF DIODE AND TRANSISTOR SWITCHES

Because of various factors, to be examined in this section, the response of a diode or transistor operating as a switch is delayed and the waveshape distorted. As a result, the rate at which a diode or transistor can be switched is limited.

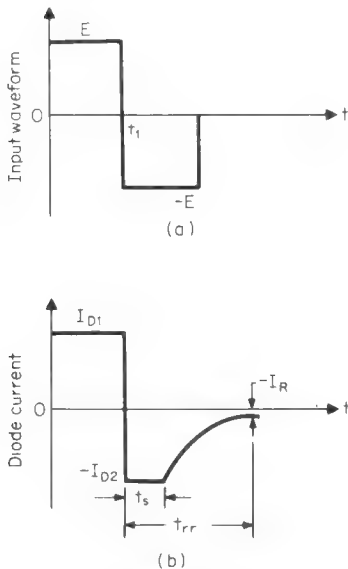
**RESPONSE OF A DIODE SWITCH** Assume that the waveform of Fig. 14.43a is applied to a diode, such as a diode in a positive OR gate. During the interval of  $t_1$  seconds, the diode is forward-biased and a current  $I_{D1}$  flows, as indicated in Fig. 14.43b.

At  $t = t_1$ , the input voltage is  $-E$  volts, and the diode becomes reverse-biased. What one expects is that the current in a reverse-biased diode should be zero. Instead, an appreciable current,  $-I_{D2}$ , flows for a time  $t_s$  seconds, which eventually goes to zero.

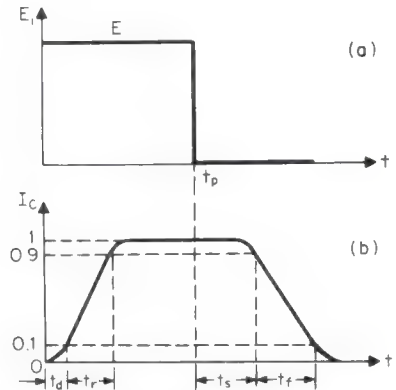
To explain this anomalous behavior, it is necessary to examine the action of a pn junction (see Chap. 8). When a diode is forward-biased, there is a huge buildup of electrons in the p region and holes in the n region in the vicinity of the pn junction. When the diode is reverse-biased, it takes time for the electrons and holes to depart from the junction region. Consequently the diode conducts in the reverse direction for a time  $t_s$ , referred to as the *storage time*.

After  $t_s$  has elapsed, the reverse current gradually decreases and ultimately reaches zero. The reason for the gradual decay in reverse current is that it takes time for the diode depletion capacitance to charge to the reverse voltage,  $-E$  volts. The time from  $t_1$ , when the diode is first reverse-biased, to the time when the reverse current equals a specified value,  $-I_R$ , is the *reverse-recovery time*  $t_{rr}$  of the diode. For good switching diodes,  $t_{rr}$  is in the order of nanoseconds ( $10^{-9}$  s).

**RESPONSE OF A TRANSISTOR SWITCH** Assume that the rectangular pulse (length =  $t_p$ ) of Fig. 14.44a is impressed across the input terminals of an INVERTER.



**Fig. 14.43** Response of a diode switch: (a) Voltage waveform applied to diode. (b) Diode current flow in response to impressed voltage waveform.



**Fig. 14.44** Response of a transistor switch: (a) Pulse applied to the input of an INVERTER. (b) Collector current waveform in response to the input pulse.

The collector current appears as shown in Fig. 14.44b. The output waveshape is stretched and distorted. For convenience, the maximum value of collector current is normalized to one. Four specific times are indicated in the figure. These are:

1. *Delay time*  $t_d$  is defined as the time it takes the collector current to reach 0.1 (10 percent) of its maximum value. It corresponds to the time required to charge the junction capacitance plus the time it takes for the carriers (electrons or holes) to reach the collector region.

2. *Rise time*  $t_r$  is defined as the time it takes the collector current to rise from 0.1 (10 percent) to 0.9 (90 percent) of its maximum value. The rise time is the time needed to charge additional transistor capacitances.

3. *Storage time*  $t_s$  is defined as the time, measured from  $t_p$  to where the collector

current falls to 90 percent of its maximum value. In saturation, both the base-emitter and collector-base junctions are forward-biased. Consequently the collector, in addition to the emitter, injects electrons (or holes) into the base region. Similar to turning off a forward-biased diode, it takes time for the carriers to leave the base region. This phenomenon results in a storage time delay.

4. *Fall time*  $t_f$  is defined as the time required for the collector current to fall from its 90 to 10 percent maximum values. During this interval, the transistor capacitances are being discharged.

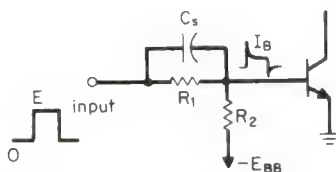
The sum of the delay and rise times,  $t_d + t_r$ , is equal to the *turn-on time*  $t_{on}$  of the transistor:

$$t_{on} = t_d + t_r \quad (14.13a)$$

Similarly, the sum of the storage and fall times,  $t_s + t_f$ , is equal to the *turn-off time*,  $t_{off}$ , of the transistor:

$$t_{off} = t_s + t_f \quad (14.13b)$$

To decrease the turn-on time, the base is driven hard. In this condition, more base current is supplied to the base than required for the transistor to saturate. Because of the excess base current the storage time, and hence the turn-off time, increases. For high-speed switching, the transistor may be operated as a nonsaturated switch. An example of a nonsaturating circuit is emitter-coupled logic (ECL), considered in Chap. 15.



**Fig. 14.45** The use of a speed-up (commutating) capacitor  $C_s$  across the series input resistor  $R_1$  to improve the switching speed of a transistor.

A method used often in discrete circuits to decrease the turn-on and turn-off times is to place a *speed-up*, or *commutating*, capacitor across the resistor in series with the transistor base (Fig. 14.45). The value of the speed-up capacitor  $C_s$  is in the order of 50 pF. Initially, when the input rises from 0 to  $E$  volts, the capacitor acts as a short circuit. Maximum base current therefore flows and the turn-on time is reduced. At the termination of the input pulse, the capacitor also acts as a short circuit and aids in removing the excess carriers from the base region. As a result, the storage and turn-off times are also reduced.



# Chapter 15

## Digital Integrated Circuits

### 15.1 INTRODUCTION

In the development of the digital integrated circuit (DIC), families of logic evolved which exhibit well-defined properties. Three important families that enjoy wide use are:

1. *Transistor-transistor logic* (TTL, T<sup>2</sup>L) is the most widely used logic. It has good speed, reasonably low power dissipation per gate, and is relatively low in cost.

2. *Emitter-coupled logic* (ECL), or *current-mode logic* (CML), is the fastest logic available today. Operating as a nonsaturated switch, it can switch frequencies as high as 500 MHz. Its power dissipation per gate, as well as its cost, is relatively high.

3. *Complementary metal-oxide semiconductor logic* (CMOS, COS/MOS) exhibits the lowest dissipation per gate. It is slower than T<sup>2</sup>L or ECL, and its cost is moderate. Whereas T<sup>2</sup>L and ECL employ the bipolar junction transistor in their circuits, CMOS uses the enhancement-type MOSFET. A comparison of the three major families of logic is provided in Table 15.1.

Owing to rapid advances in integrated circuit processing, logic circuits are available with an excess of 1000 logic gates on a single chip of silicon. One can define four levels of complexity in terms of the number of gates on a chip:

1. *Small-scale integration* (SSI): A DIC containing less than 12 gates on a chip

2. *Medium-scale integration* (MSI): A DIC containing more than 12, but less than 100 gates on a chip

3. *Large-scale integration* (LSI): A DIC containing more than 100, but less than 1000 gates on a chip

4. *Grand-scale integration* (GSI): A DIC containing more than 1000 gates on a chip

This chapter defines the terms used to characterize a DIC; discusses the operation of T<sup>2</sup>L, ECL, and CMOS logic; and considers their application as flipflops in shift registers and counters. The chapter concludes with a discussion of semiconductor memories, digital-to-analog and analog-to-digital converters, and the microprocessor.

### 15.2 DIC TERMS AND PARAMETERS

In addition to the specification of required power-supply voltages and operating temperature range, a number of terms and parameters are used in characterizing the performance of a digital integrated circuit. Important specifications include:

**FAN IN** The maximum number of inputs to a gate.



### TABLE 15.1 Comparison of IC Logic Families

Parameter	TTL			ECL					
	Standard	Low-power	High-speed	Schottky	1-nS	2-nS	4-nS	8-nS	CMOS
Logic function	NAND	NAND	NAND	NAND	OR/NOR	OR/NOR	OR/NOH	OR/NOH	NOR/NAND
Supply voltage, V	5	5	5	5	-5.2	-5.2	-5.2	-5.2	3 to 18
Power dissipated/gate, mW	12	1	22	18	60	25	22	31	10-50
Propagation delay/gate, nS	10	33	6	3	1	2	4	8	70
Speed, MHz	35	3	50	125	500	200	165	30	5
Noise margin	Very good	Very good	Very good	Good	Fair	Fair	Fair	Fair	Excellent
Fan out (typical)	10	10	10	10	10	25	25	25	>50

• This is the *static* value. In switching, the power dissipation per gate at 1 MHz is approximately 1 mW and increases with frequency.

**FAN OUT** The maximum number of gates that may be connected to the output terminal of a single gate.

**PROPAGATION DELAY** The propagation delay  $t_{dp}$  is the difference in time between the application of an input signal and its presence at the output of a DIC. Its unit is generally nanoseconds.

**SPEED** This parameter indicates how fast a flipflop can change states. Its unit is megahertz.

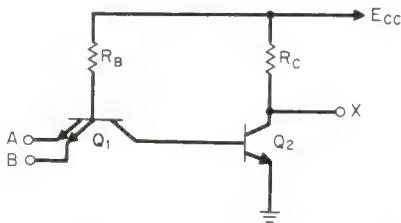
**GATE DISSIPATION** The average dc power dissipated in a gate. Its unit is milliwatts or microwatts.

**SPEED-POWER PRODUCT** The product of propagation delay (in nanoseconds) and gate dissipation (in milliwatts). Because the product of power and time yields energy, the unit for the speed-power product is picojoules, pJ ( $1 \text{ pJ} = 10^{-12} \text{ J}$ ).

**NOISE MARGIN** Noise margin NM is the maximum extraneous voltage that causes a gate to change its state. Its unit is volts or millivolts.

### 15.3 T<sup>2</sup>L

A basic T<sup>2</sup>L circuit having a fan-in of two is illustrated in Fig. 15.1. Transistor  $Q_1$  is unique; it has multiple emitters. Each base-emitter junction behaves like a diode.



**Fig. 15.1** Basic T<sup>2</sup>L gate shown with a fan-in of two. The circuit performs positive NAND logic.

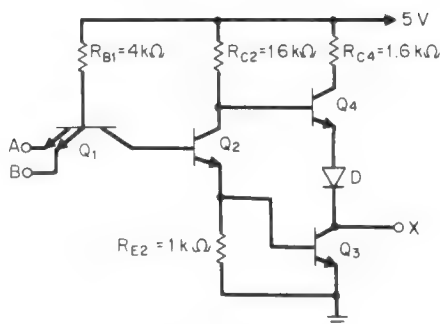
The two base-emitter junctions, therefore, act like diodes of an AND gate to which are connected inputs  $A$  and  $B$ . The base-collector junction of  $Q_1$  serves as a diode in series with the base of  $Q_2$ , which is connected as an INVERTER.

The logic performed by T<sup>2</sup>L is positive NAND logic. Assume that either input  $A$  or  $B$  is at logic 0. One of the base-emitter junctions becomes forward-biased. The base of  $Q_1$  is therefore near ground, and the collector-base junction is reverse-biased. No base current flows in  $Q_2$ , and the transistor is OFF:  $X = 1$ .

If both inputs are at logic 0, the same condition exists as though either one is at logic 0 and output  $X = 1$ . When both  $A$  and  $B$  are at logic 1, the base-emitter diodes are reverse-biased. The collector-base junction of  $Q_1$  is now forward-biased, and base current flows in  $Q_2$ . Output  $X$ , therefore, is at logic 0.

There exists a number of members in the T<sup>2</sup>L family. They include the standard, low-power, high-speed, and Schottky-clamped series. These circuits are examined in the following discussion.

**STANDARD T<sup>2</sup>L** The circuit for a two-input standard T<sup>2</sup>L NAND gate is shown in Fig. 15.2. Transistor  $Q_1$  serves the same function as in the basic circuit of Fig. 15.1. Transistor  $Q_2$  is connected as a *phase splitter*. When a logic 0 is at its base, its collector is at a logic 1 and its emitter at a logic 0. If the input is equal to a logic 1, the reverse is true. Now, the collector is at a logic 0 and the emitter at a logic 1.



**Fig. 15.2** Circuit for a two-input standard T<sup>2</sup>L NAND gate.

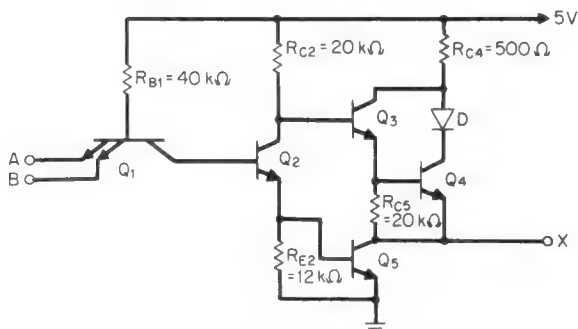
Transistors  $Q_3$  and  $Q_4$  are connected as a *totem-pole amplifier*. This configuration minimizes the effects of any capacitance present, such as stray, across the output. One serious effect of capacitance across the output of a gate is to reduce its switching speed. The function of diode  $D$  is to limit the collector current to a reasonable value when  $Q_3$  is in the ON state.

Assume that inputs  $A$  and  $B$  are at logic 0. The base-emitter junctions of  $Q_1$  are forward-biased, and a logic 0 appears at the base of  $Q_2$ . Owing to the operation of the phase splitter, at the collector of  $Q_2$  there appears a logic 1 and at its emitter a logic 0. Transistor  $Q_4$  is therefore ON and  $Q_3$  is OFF. Output  $X$  is at a potential equal to  $E_{CC}$  less the voltage drops across  $R_{C4}$ , the collector-emitter of  $Q_4$ , and the diode; hence,  $X = 1$ .

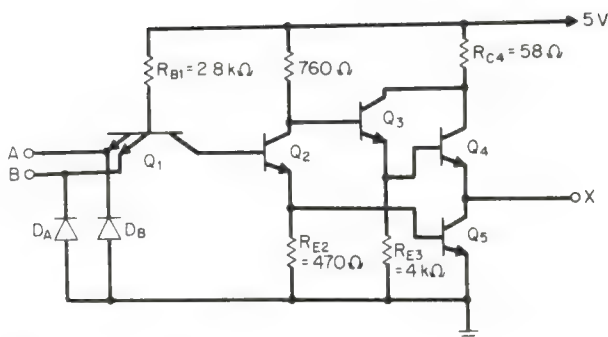
If either input is at logic 0, the same condition described in the preceding paragraph prevails, and  $X = 1$ . Assume that both inputs are at logic 1. A logic 1 therefore appears at the base of  $Q_2$ . As a result, a logic 0 is at the collector and a logic 1 at the emitter of  $Q_2$ . Transistor  $Q_4$  is now OFF, and  $Q_3$  is ON;  $X = 0$ .

**LOW-POWER T<sup>2</sup>L** A version of T<sup>2</sup>L which dissipates less than one-tenth the power of standard T<sup>2</sup>L is illustrated in Fig. 15.3. Referred to as low-power T<sup>2</sup>L, the NAND gate is very similar to the standard circuit. To ensure low-power dissipation, the resistance values are increased appreciably. For example,  $R_{B1} = 40 \text{ k}\Omega$ , and  $R_{C2} = 20 \text{ k}\Omega$ ; in the standard T<sup>2</sup>L, their corresponding values are 4 and 1.6  $\text{k}\Omega$ , respectively.

Transistors  $Q_3$  and  $Q_4$  constitute a Darlington pair. In addition to providing greater current gain than a single transistor, the Darlington pair also increases the switching speed of the gate.



**Fig. 15.3** An example of a low-power T<sup>2</sup>L gate. Transistors  $Q_3$  and  $Q_4$  are connected as a Darlington pair which provides more current gain than a single transistor.



**Fig. 15.4** High-speed T<sup>2</sup>L gate. It features low values of resistances. Diodes  $D_A$  and  $D_B$  protect the base-emitter junctions of transistor  $Q_1$  against voltage breakdown.

**HIGH-SPEED T<sup>2</sup>L** Similar in configuration to low-power T<sup>2</sup>L, a high-speed T<sup>2</sup>L NAND gate is shown in Fig. 15.4. To achieve high-speed operation, the resistance values in the circuit are kept low. Because at high switching speeds stray inductance, in addition to stray capacitance, enters the picture, oscillations may be superimposed on the pulses. This phenomenon is referred to as *ringing*. The purpose of diodes  $D_A$  and  $D_B$ , referred to as *clamping diodes*, connected between each input and ground is to limit negative signal swings because of ringing. In this manner the base-emitter junctions are protected from avalanche or zener breakdown.



**Fig. 15.5** Schottky-clamped transistor: (a) Schottky diode connected across the base and collector terminals of a junction transistor. (b) Symbol for a Schottky-clamped transistor, referred to as a Schottky transistor.

**SCHOTTKY-CLAMPED T<sup>2</sup>L** The transistors in the previous examples of transistor-transistor logic all operate in saturation when in the ON state. As explained in Chap. 14, a saturated switch gives rise to storage time which reduces its speed of operation. To increase switching speed, the transistor is operated as a nonsaturated switch. In this mode of operation, the transistor in the ON state is in the active region close to, but never in, saturation.

To achieve efficient nonsaturated switching, a Schottky diode (see Chap. 8) is connected across the collector and base of a transistor, as illustrated in Fig. 15.5a. When the transistor is turned on, some of the base current is diverted by the diode from the base. Consequently, less base current flows, and the transistor is prevented from saturating. The symbol for a Schottky-clamped transistor, also referred to as a *Schottky transistor*, is given in Fig. 15.5b.

An example of a Schottky-clamped T<sup>2</sup>L NAND gate is illustrated in Fig. 15.6. The circuit is similar to the high-speed T<sup>2</sup>L gate of Fig. 15.4. A low-power Schottky-clamped T<sup>2</sup>L gate is also available. As in the low-power T<sup>2</sup>L gate, high-resistance values are used in this version.

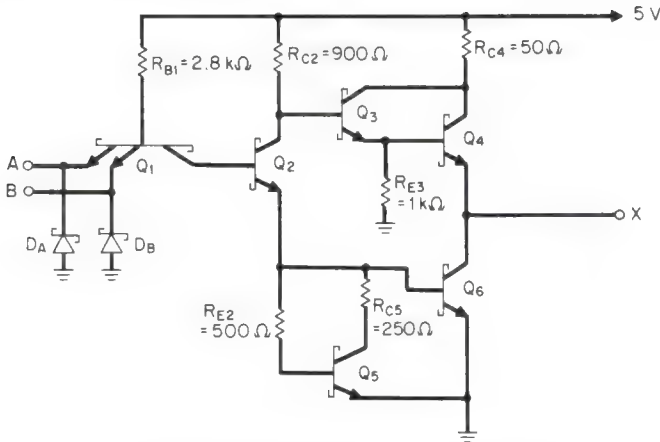


Fig. 15.6 An example of a Schottky-clamped T<sup>2</sup>L gate.

#### 15.4 ECL

With its transistors operating as nonsaturated switches, emitter-coupled logic (ECL), also referred to as current-mode logic (CML), is the fastest family of logic circuits. It is available in four basic types having propagation delays of 8, 4, 2, and 1 ns (see Table 15.1). Emitter-coupled logic provides two logic functions: the NOR and OR. Its logic symbol is illustrated in Fig. 15.7.

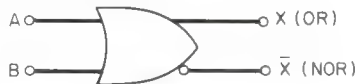


Fig. 15.7 Logic symbol for an ECL gate.

A recent member of the family is the 2-nS ECL gate. This version optimizes switching speed and gate dissipation (see Table 15.1). An example of a 2-nS gate having a fan-in of two is provided in Fig. 15.8. It is powered by a  $-5.2\text{-V}$  dc power supply. A logic 0 is approximately  $-1.7\text{ V}$ , and a logic 1 is equal to  $-0.9\text{ V}$ . Because a logic 1 is less negative than a logic 0, their values correspond to positive logic.

A basic circuit of ECL is the differential amplifier comprised of transistors  $Q_2$  and  $Q_3$ .

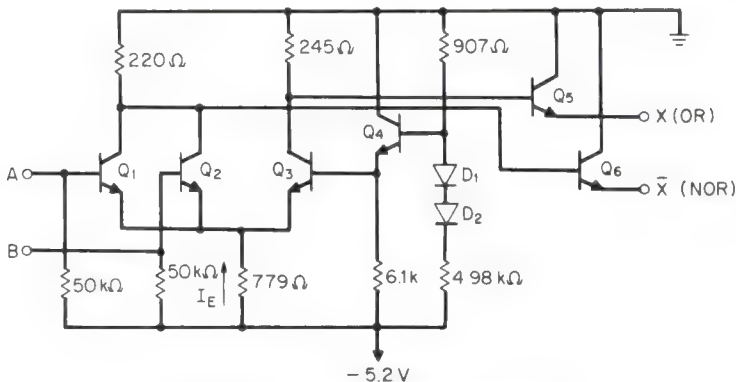


Fig. 15.8 An example of a 2-nS ECL gate.

(see Chap. 13). Emitter current  $I_E$  is essentially constant. The base of  $Q_3$  is held at a constant dc reference voltage of 1.29 V. The reference voltage is stabilized with respect to temperature by transistor  $Q_4$  and diodes  $D_1$  and  $D_2$ . Switching occurs when the input to either transistor,  $Q_1$  or  $Q_2$ , is approximately 0.1 V greater or less than the reference voltage. Emitter followers  $Q_5$  and  $Q_6$  provide dc level shifting to ensure that a logic 0 corresponds to  $-1.7$  V and a logic 1 to  $-0.9$  V.

Assume that inputs  $A$  and  $B$  are at logic 0. For this condition,  $Q_1$  and  $Q_2$  are OFF and  $Q_3$  is ON (logic 0). The output of  $Q_5$ , therefore, is  $X = 0$  (OR) and at  $Q_6$ ,  $\bar{X} = 1$  (NOR).

Suppose that  $A$  is at logic 0 and  $B$  at logic 1. Now  $Q_3$  is OFF,  $Q_2$  is ON; and  $X = 1$ ,  $\bar{X} = 0$ . If both inputs are at logic 1, as in the preceding case,  $X = 1$  and  $\bar{X} = 0$ . We see therefore that ECL provides an OR and a NOR output.

## 15.5 CMOS

Complementary metal-oxide semiconductor (CMOS) logic dissipates minute power and can operate over a power-supply range of 3 to 18 V. It is, however, slower than T<sup>2</sup>L or ECL (see Table 15.1). In this logic family, no resistors are used. Instead, p-channel and n-channel enhancement-type MOSFETs are employed in complementary pairs. As a result, a much greater packing density is realized than with the bipolar junction transistor in monolithic integrated circuits.

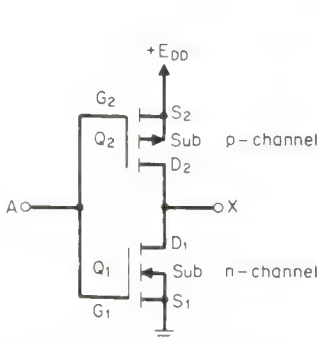
**INVERTER** A CMOS inverter is illustrated in Fig. 15.9. Source  $S_2$  of the p-channel device is returned to the supply voltage ( $E_{DD}$  volts), and source  $S_1$  of the n-channel device is returned to ground. Output  $X$  is obtained at the junction of the drain terminals,  $D_1$  and  $D_2$ . Both gates, which are connected together, constitute input terminal  $A$ . The substrate (Sub) is connected to the source of each transistor.

A logic 1 is equal to  $E_{DD}$  volts and a logic 0 to 0 V. Assume that  $A$  is at logic 1. Because  $G_1$  is positive with respect to  $S_1$ , the n-channel MOSFET ( $Q_1$ ) is ON. The gate-source voltage of the p-channel device ( $Q_2$ ), however, is at 0 volts ( $E_{DD} - E_{DD} = 0$ ); hence,  $Q_2$  is OFF. Output  $X = 0$ .

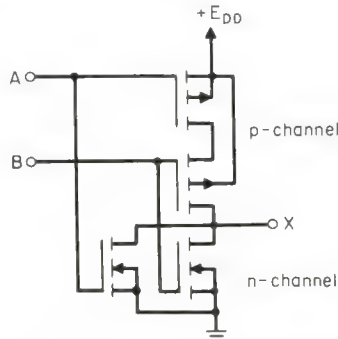
If  $A$  is at logic 0, the n-channel transistor is OFF. The p-channel device, whose gate is now negative with respect to its source ( $0 - E_{DD} = -E_{DD}$ ), is ON. The output in this condition may be thought of as being shifted to  $E_{DD}$  volts and  $X = 1$ .

In either case, one transistor is always ON and the other is always OFF. Since the transistors are connected in series, the drain current flowing is equal to their leakage current. Because of its extremely small value, the dissipation is in the order of 0.01  $\mu$ W. (In switching, however, the gate dissipation is approximately 1 mW at 1 MHz and increases with frequency.)

**NOR GATE** An example of a two-input NOR gate employing CMOS logic is shown in Fig. 15.10. Two p-channel MOSFETs are connected in series and two n-channel



**Fig. 15.9** A CMOS INVERTER. Transistor  $Q_1$  is an n-channel enhancement-type MOSFET and  $Q_2$  a p-channel device.



**Fig. 15.10** A CMOS two-input NOR gate.



## 15-8 Digital Integrated Circuits

MOSFETs are connected in parallel. The gate of each n-channel transistor is connected to the gate of a p-channel transistor. The substrates of the p-channel devices are connected to  $E_{DD}$  and the substrates of the n-channel devices are returned to ground.

Assume that  $A$  and  $B$  are at logic 0. In this case both p-channel devices are ON, and the n-channel devices are OFF. Output  $X = 1$  (the NOT of a zero is one). If  $A$  or  $B$ , or both, are at logic 1, one or both p-channel transistors are OFF. Also, one or both n-channel transistors are ON. Output  $X = 0$  (the NOT of a one is zero).

**example 15.1** Show how the two-input NOR gate of Fig. 15.10 may be expanded into a three-input NOR gate.

**solution** This is accomplished by adding a p-channel MOSFET in series and an n-channel MOSFET in parallel (Fig. 15.11). Their gates are connected together for the third input  $C$ . In such fashion, one may develop a four-, or greater, input NOR gate.

**NAND GATE** A two-input CMOS NAND gate is illustrated in Fig. 15.12. In contrast with the NOR gate of Fig. 15.10, the p-channel devices are in parallel and the n-channel devices are in series in the NAND gate. For each additional input, a p-channel MOSFET is connected in parallel and an n-channel MOSFET in series. Their gates are then joined together for the new input.

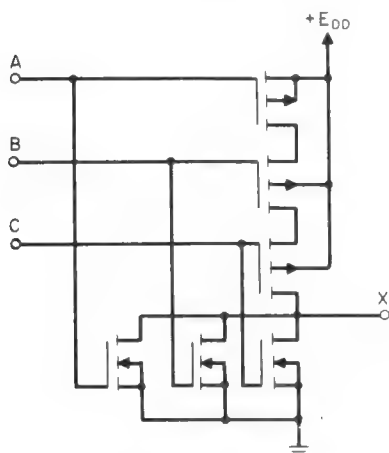


Fig. 15.11 A CMOS three-input NOR gate. (See Example 15.1.)

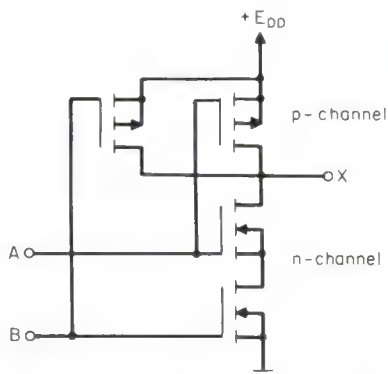


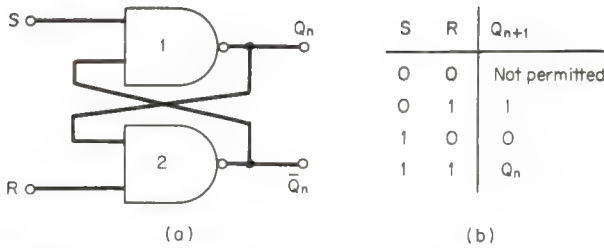
Fig. 15.12 A CMOS two-input NAND gate.

Assume that both  $A$  and  $B$  are at logic 1. The p-channel devices are OFF, and the n-channel devices are ON. Output  $X = 0$  (the NOT of a one is zero). If either or both  $A$  and  $B$  are at logic 0, one or both p-channel transistors are ON and one or both n-channel transistors are OFF. Output  $X = 1$  (the NOT of a zero is one).

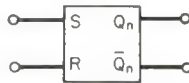
## 15.6 FLIPFLOPS

The bistable MV, described in Chap. 14, is the basic circuit in a number of flipflops used in shift registers and counters. In this section, their behavior is illustrated by truth tables. Because the gates of logic families are generally either NAND or NOR (see Table 15.1), the implementation of flipflops is based on these configurations.

**R-S FLIPFLOP (LATCH)** An R-S (reset-set) flipflop, also referred to as a *latch*, implemented with NAND gates is illustrated in Fig. 15.13a. The output of each gate is connected to the input of the other gate. In the truth table of Fig. 15.13b, subscript  $n + 1$  indicates the state of the flipflop after being triggered by a suitable pulse at either the  $R$  or  $S$  input terminals. Subscript  $n$  denotes the state of the flipflop before being triggered. The logic symbol for an R-S flipflop is provided in Fig. 15.14.



**Fig. 15.13** An R-S flipflop (latch) using NAND gates: (a) Circuit. (b) Truth table.



**Fig. 15.14** Logic symbol for an R-S flipflop.

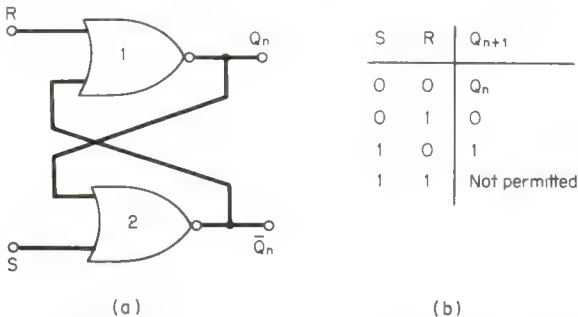
For no input trigger ( $S = R = 0$ ), both outputs tend to go to a logic 1 (the NOT of a zero is one). Because this is an unstable state, this condition is not permitted.

Assume that  $Q_n = 0$  and, therefore,  $\bar{Q}_n = 1$ . Because output  $\bar{Q}_n$  is an input to NAND gate 1, one of its inputs, therefore, is a logic 1. If  $S = 0$ , the output of gate 1 is a one (the NOT of a zero is a one); hence,  $Q_{n+1} = 1$ . Similarly, if  $Q_n = 1$  and  $R = 0$ ,  $Q_{n+1}$  becomes a 0. For  $S = R = 1$ , the state of the flipflop is unchanged.

**example 15.2** Show how an R-S flipflop may be implemented using NOR gates. Draw a truth table for the flipflop.

**solution** The circuit is shown in Fig. 15.15a and the truth table in Fig. 15.15b. For  $S = R = 0$ , there is no change in state ( $Q_{n+1} = Q_n$ ). Assume that  $Q_n = 1$ ; therefore  $\bar{Q}_n = 0$ . Output  $\bar{Q}_n$  is one of the inputs to NOR gate 1. If  $R = 1$ , the NOT of  $1 + 0 = 0$ ; hence  $Q_{n+1} = 0$  and  $\bar{Q}_{n+1} = 1$ .

Now assume that  $Q_n = 0$  ( $\bar{Q}_n = 1$ ). Output  $Q_n$  is one of the inputs to NOR gate 2. If  $S = 1$ , the NOT of  $0 + 1 = 0$ ; hence  $Q_{n+1} = 1$  and  $\bar{Q}_{n+1} = 0$ . The condition  $S = R = 1$  is not permitted.

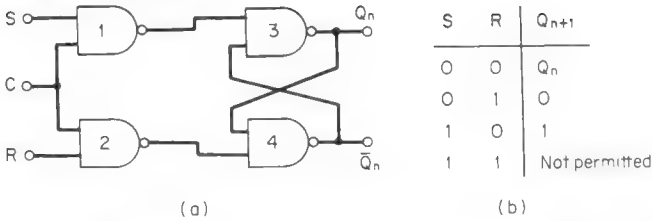


**Fig. 15.15** An R-S flipflop using NOR gates: (a) Circuit. (b) Truth table. (See Example 15.2.)

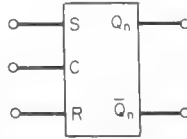
**CLOCKED R-S FLIPFLOP** In many applications, such as shift registers, it is necessary that the operation of flipflops be synchronized by a master clock in the system. Such a flipflop is referred to as a clocked or *synchronous* flipflop. An example of a clocked R-S flipflop employing NAND gates and its truth table are illustrated in Fig. 15.16. The logic symbol for a clocked R-S flipflop is shown in Fig. 15.17.

NAND gates 1 and 2 are referred to as *steering gates*. A clock pulse  $C$  is one of the inputs to each steering gate. NAND gates 3 and 4 constitute the R-S flipflop of Fig.

## 15-10 Digital Integrated Circuits



**Fig. 15.16** A clocked R-S flipflop using NAND gates: (a) Circuit. (b) Truth table.



**Fig. 15.17** Logic symbol for a clocked R-S flipflop.

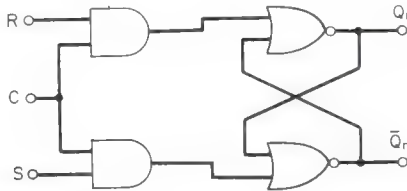
15.13a. If  $S = R = 0$ , the flipflop remains in its original state. The condition  $S = R = 1$  is not permitted.

Assume that  $Q_n = 0$  ( $\bar{Q}_n = 1$ ). A trigger at terminal  $S$  and a clock pulse to gate 1 result in a zero at its output. At the input of gate 3, therefore, there is a 0 from gate 1, and a 1 owing to the output of gate 4. The output of gate 3 is now a logic 1 ( $Q_{n+1} = 1$ ). The flipflop has changed its state.

To return the flipflop to its original state ( $Q_n = 0$ ), a trigger is applied to terminal  $R$  and a clock pulse to the other terminal of gate 2. The inputs to gate 4 are now a 0 from gate 2 and a 1 owing to the output of gate 3. Hence,  $Q_{n+1} = 0$  and  $\bar{Q}_{n+1} = 1$ .

**example 15.3** Show how a clocked R-S flipflop may be implemented using NOR and AND gates. Draw a truth table for the flipflop.

**solution** The circuit is given in Fig. 15.18; the truth table is identical with that of Fig. 15.16b.



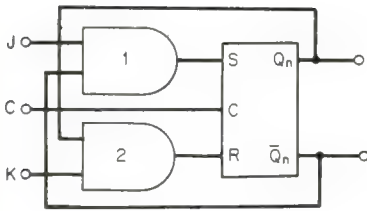
**Fig. 15.18** A clocked R-S flipflop employing NOR and AND gates. (See Example 15.3.)

**J-K FLIPFLOP** The circuit of a clocked J-K flipflop and its truth table are provided in Fig. 15.19. The input terminals, in addition to the  $C$  terminal, are denoted by  $J$  and  $K$ . The basic circuit is the clocked R-S flipflop considered in the preceding section. The outputs of AND gates 1 and 2 are connected to the  $S$  and  $R$  terminals, respectively.

The first three entries in the truth table of Fig. 15.19b are identical with the clocked R-S flipflop. Although  $S = R = 1$  is not permitted in the R-S flipflop,  $J = K = 1$  is allowed in the J-K flipflop. An examination of the truth table reveals that for this input condition the output is complemented ( $\bar{Q}_n$ ). That is, if  $Q_n = 1$ ,  $Q_{n+1} = 0$  and if  $Q_n = 0$ ,  $Q_{n+1} = 1$ . The logic symbol for the J-K flipflop is given in Fig. 15.20.

**example 15.4** Show how a J-K flipflop may be implemented using NAND gates.

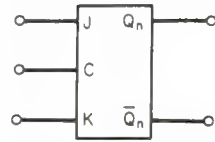
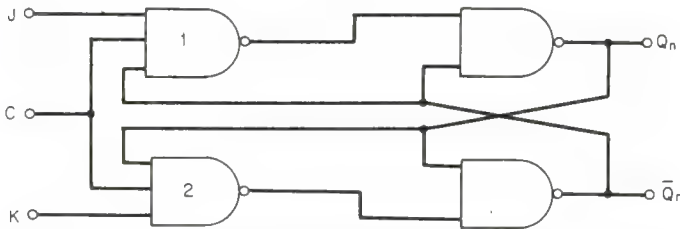
**solution** This is accomplished by replacing the two-input NAND gates (1 and 2) of Fig. 15.16a by three-input NAND gates, as illustrated in Fig. 15.21.



(a)

J	K	$Q_{n+1}$
0	0	$Q_n$
0	1	0
1	0	1
1	1	$\bar{Q}_n$

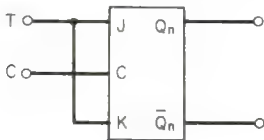
(b)

**Fig. 15.19** The *J-K* flipflop: (a) Circuit. (b) Truth table.

**Fig. 15.20** Logic symbol for a *J-K* flipflop.

**Fig. 15.21** A *J-K* flipflop implemented with NAND gates. (See Example 15.4.)

**T FLIPFLOP** This flipflop changes its state with each application of a trigger to a single input terminal *T*. Also referred to as a *toggle*, the *T* flipflop is used often in counters. An example of a clocked *T* flipflop implemented by using a *J-K* flipflop, and its truth table, are provided in Fig. 15.22. It is seen that the *T* input terminal is formed by joining together the *J* and *K* terminals. The logic symbol for the *T* flipflop is shown in Fig. 15.23.

**example 15.5** Show how a *T* flipflop may be realized using AND gates and an *R-S* flipflop.

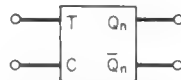
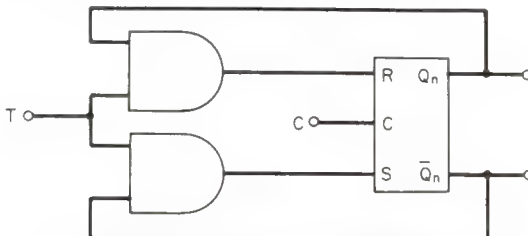
**solution** The circuit is given in Fig. 15.24.



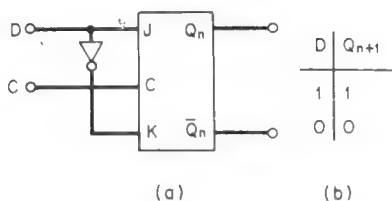
(a)

T	$Q_{n+1}$
1	$\bar{Q}_n$
0	$Q_n$

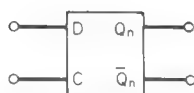
(b)

**Fig. 15.22** A *T* (toggle) flipflop implemented with a *J-K* flipflop: (a) Circuit. (b) Truth table.

**Fig. 15.23** Logic symbol for a *T* flipflop.

**Fig. 15.24** A *T* flipflop realized by using AND gates and an *R-S* flipflop. (See Example 15.5.)

## 15-12 Digital Integrated Circuits



**Fig. 15.25** A delay ( $D$ ) flipflop implemented with a  $J$ - $K$  flipflop and an INVERTER: (a) Circuit. (b) Truth table.



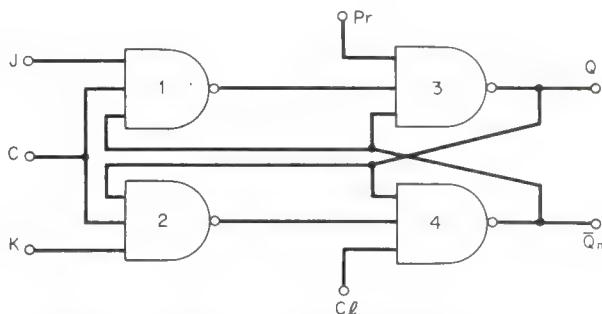
**Fig. 15.26** Logic symbol for a  $D$  flipflop.

**D FLIPFLOP** An example of a clocked  $D$ , or *delay*, flipflop using a  $J$ - $K$  flipflop and an INVERTER is given in Fig. 15.25a. Its truth table is provided in Fig. 15.25b. An  $R$ - $S$  flipflop may be used instead of the  $J$ - $K$  type. In this case the input to the INVERTER is connected to the  $S$  terminal and the output to the  $R$  terminal. The  $D$  terminal is also connected to  $S$ . The logic symbol for the  $D$  flipflop is shown in Fig. 15.26.

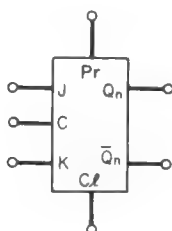
The  $D$  flipflop is used as a one-bit delay. When  $D = 1$ ,  $J = 1$  and, because of the INVERTER,  $K = 0$ . Referring to the truth table of Fig. 15.19b for the  $J$ - $K$  flipflop,  $Q_{n+1} = 1$ . Also, if  $D = 0$ ,  $J = 0$ , and  $K = 1$ ; hence, from the truth table,  $Q_{n+1} = 0$ . Output  $Q_{n+1}$  therefore equals  $D$  delayed by one clock pulse, or bit.

**CLEAR AND PRESET** It is often necessary to *clear* a flipflop to its zero ( $Q = 0$ ) or *preset* it to its one ( $Q = 1$ ) state. An example of how this is realized is illustrated in Fig. 15.27. All four NAND gates have three inputs each. One input to gate 3 is the *preset*,  $Pr$ , input; and  $Cl$  to gate 4 is the *clear* input. The logic symbol for a clear-preset  $J$ - $K$  flipflop is illustrated in Fig. 15.28.

In either clearing or presetting a flipflop, the clock pulse is at logic zero ( $C = 0$ ). Assume that the flipflop is to be preset to one ( $Q = 1$ ). Inputs  $Pr = 0$ ,  $Cl = 1$  and, as mentioned previously,  $C = 0$ . Because  $Pr = 0$ , even if the other two inputs to gate 3 are



**Fig. 15.27** Adding clear ( $Cl$ ) and preset ( $Pr$ ) inputs to a  $J$ - $K$  flipflop.



**Fig. 15.28** Logic symbol for a clear-preset  $J$ - $K$  flipflop.





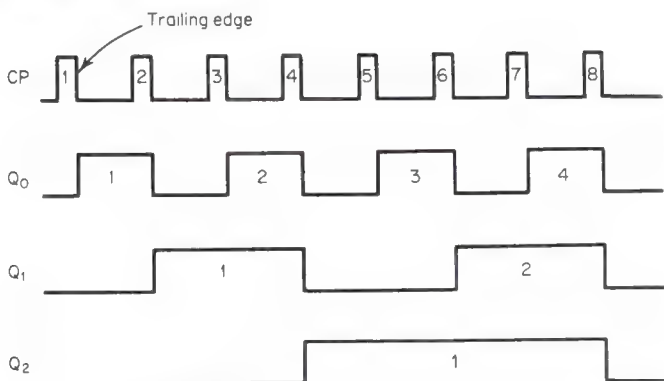


Fig. 15.31 Timing diagram for the ripple counter of Fig. 15.30.

example, a three-stage ripple counter implemented with *T* flipflops is shown. The *J* and *K* terminals of each flipflop are connected and returned to a logic 1. This connection converts the *J-K* to a *T* flipflop (see Fig. 15.22a). Because the *J-K* terminals are returned to a logic 1, output *Q* of each flipflop is at logic zero. The outputs of *FF*<sub>0</sub>, *FF*<sub>1</sub>, and *FF*<sub>2</sub> are designated as *Q*<sub>0</sub>, *Q*<sub>1</sub>, and *Q*<sub>2</sub>, respectively. Output *Q*<sub>0</sub> is the LSB and *Q*<sub>2</sub> the MSB.

Assume that each flipflop changes its state on the falling (trailing), or *negative-going edge*, of a pulse. A timing diagram for the ripple counter of Fig. 15.30 is given in Fig. 15.31. On the falling edge of *CP*<sub>1</sub>, *Q*<sub>0</sub> goes to a logic 1. Because each flipflop changes its state on the falling edge, *FF*<sub>1</sub> and *FF*<sub>2</sub> are unaffected; that is, *Q*<sub>1</sub> = *Q*<sub>2</sub> = 0. Thus, after one clock pulse, we have a binary count of 1:

$$\begin{array}{r} Q_2 \ Q_1 \ Q_0 \\ 0 \ 0 \ 1 \end{array}$$

At the end of *CP*<sub>2</sub>, *Q*<sub>0</sub> in going to zero triggers *Q*<sub>1</sub>; *Q*<sub>2</sub> remains unchanged. Hence, we now have a binary count of 2:

$$\begin{array}{r} Q_2 \ Q_1 \ Q_0 \\ 0 \ 1 \ 0 \end{array}$$

Proceeding in this manner, at the end of *CP*<sub>7</sub>, we have a binary count of 7:

$$\begin{array}{r} Q_2 \ Q_1 \ Q_0 \\ 1 \ 1 \ 1 \end{array}$$

At the end of *CP*<sub>8</sub>, the counter resets itself: that is, *Q*<sub>0</sub> = *Q*<sub>1</sub> = *Q*<sub>2</sub> = 0. For a four-stage counter, the counter resets at the end of the 16th pulse, and so on.

It is interesting to note from the timing diagram that the output of *FF*<sub>0</sub> changes its state once for every two clock pulses; the output of *FF*<sub>1</sub> changes its state once for every four clock pulses, and so on. A ripple counter, therefore, may be used as a *divide-by 2, 4, 8, etc.*, circuit.

**OTHER TYPES OF COUNTERS** The ripple counter of Fig. 15.30 is an example of an *asynchronous counter*. In an asynchronous counter, the individual flipflops are not clocked simultaneously. In a *synchronous counter*, all flipflops are clocked and change their states simultaneously. Definitions of other counters include:

**Up counter** An up, or *forward*, counter adds each input pulse to the count. The ripple counter of Fig. 15.30 is an example of an up counter.

**Down counter** A down, or *reverse*, counter subtracts 1 from a preset number during each input pulse.

**Up-down counter** An up-down, or *bidirectional*, counter can operate as either an up or a down counter.

**Modulo counter** A counter that counts other than binary multiples is referred to as a

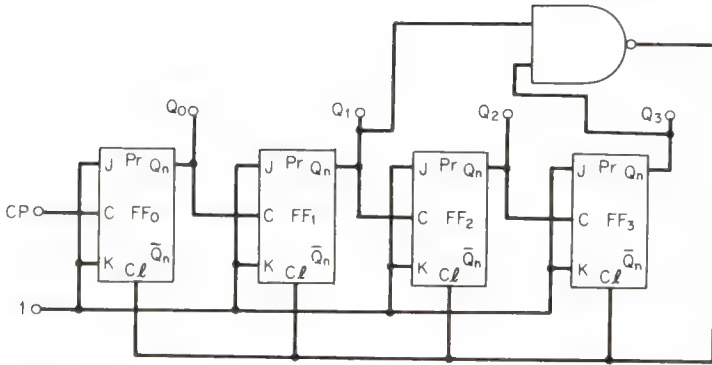


Fig. 15.32 An example of a decade counter using *J-K* flipflops. (See Example 15.6.)

modulo counter. For example, a counter that counts 1, 2, 3, 4 and resets itself to 0 on the 5th pulse is a modulo-5 counter. A *decade counter* that counts 1, . . . , 9 and resets itself to 0 on the 10th pulse is a modulo-10 counter.

**example 15.6** Using *J-K* flipflops, design a decade counter.

**solution** A decade (modulo-10) counter employing *J-K* flipflops is illustrated in Fig. 15.32. Four flipflops are required. If three flipflops are used, a maximum count of only 7 (111) is possible, which is inadequate for a decade counter.

On the 10th input pulse, the decade counter must reset to zero. Decimal 10 equals binary 1010. Therefore, on the 10th pulse,  $Q_3 = 1$  (MSB),  $Q_2 = 0$ ,  $Q_1 = 1$ , and  $Q_0 = 0$  (LSB). Outputs  $Q_3$  and  $Q_1$  are fed to the two-input NAND gate. When, on the count of 10,  $Q_1 = Q_3 = 1$ , the output of the NAND gate equals zero (the NOT of 1 and 1 = 0). The zero is returned to the clear (Cl) inputs of each flipflop, thereby resetting the counter to zero.

## 15.8 SHIFT REGISTERS

A shift register may be regarded as a component for storing a binary number or an instruction. For example, if a binary number or instruction has  $n$  bits, it is referred to as an  $n$ -bit word. An  $n$ -bit shift register consists of  $n$ -cascaded flipflops, indicated symbolically in Fig. 15.33. Flipflop  $FF_0$  contains the least significant bit (LSB) and  $FF_{n-1}$  the most significant bit (MSB) of the word stored in the register.



Fig. 15.33 A symbolic representation of an  $n$ -bit shift register.

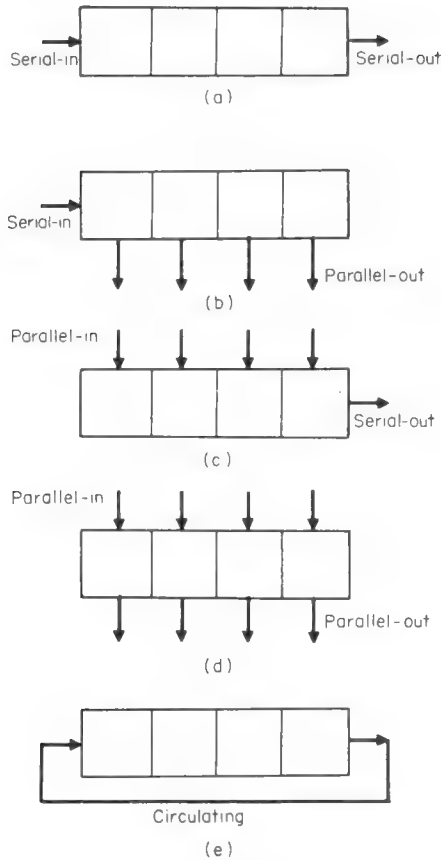
There are a number of varieties of shift registers, as illustrated symbolically in Fig. 15.34. For simplicity, four-bit registers are shown.

**Serial register** (Fig. 15.34a) In this register the word is stored and read out serially, that is, in sequence. Because the word is shifted to the *right*, it is referred to as a *right-shift* register. A register that shifts to the *left* is called a *left-shift* register.

**Serial-in, parallel-out register** (Fig. 15.34b) In this register, the word is stored serially and read out in parallel. The serial-in, parallel-out register is also referred to as a *serial-to-parallel converter*.

**Parallel-in, serial-out register** (Fig. 15.34c) In this register the word is stored in parallel (each bit being stored simultaneously) and read out serially. The parallel-in, serial-out register is also referred to as a *parallel-to-serial converter*.

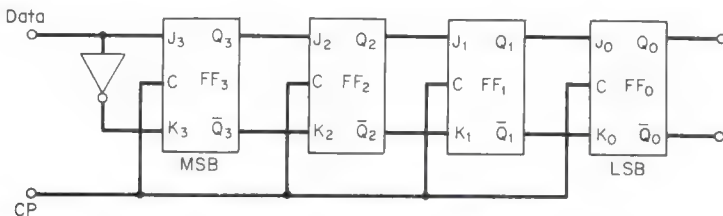
**Parallel-in, parallel-out register** (Fig. 15.34d) In this register the word is stored and read out in parallel.



**Fig. 15.34** Different types of shift registers. For simplicity, four-bit registers are illustrated: (a) Serial. (b) Serial-in, parallel-out (serial-to-parallel converter). (c) Parallel-in, serial-out (parallel-to-serial converter). (d) Parallel-in, parallel-out. (e) Circulating (dynamic shift register; shift-right read-only memory).

**Circulating register** (Fig. 15.34e) In this register the word circulates continuously. The circulating register is also referred to as a *dynamic shift register* and a *shift-right read-only memory*.

An example of a four-bit serial shift register employing *J-K* flipflops is illustrated in Fig. 15.35. During a clock pulse, each flipflop assumes the state of its preceding flip-



**Fig. 15.35** An example of a four-bit serial shift register employing *J-K* flipflops.

flop. Assume that the word to be stored in the register is 1011. The sequence of operations is illustrated in Table 15.2.

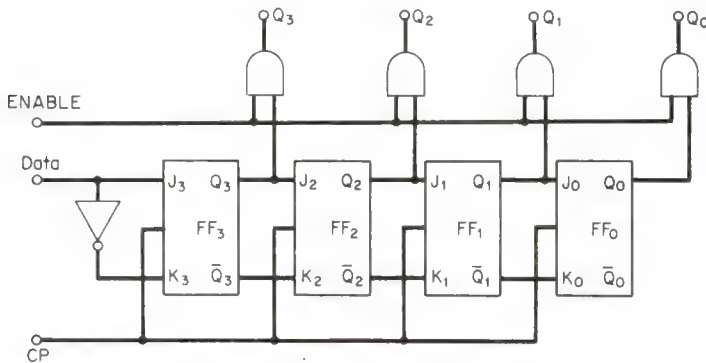
Before  $CP_1$ , the state of each flipflop is at a logic zero ( $Q_3 = Q_2 = Q_1 = Q_0 = 0$ ). At the termination of  $CP_1$ , the LSB, 1, is stored in  $FF_3$ ; hence,  $Q_3 = 1$  and  $Q_2 = Q_1 = Q_0 = 0$ . At the end of  $CP_2$ , the LSB has been shifted to  $FF_2$  and the next bit, 1, is shifted to  $FF_3$ ; hence,  $Q_3 = Q_2 = 1$  and  $Q_1 = Q_0 = 0$ .

**TABLE 15.2 Storing the Word 1011 in a Serial Shift Register**

CP	Bit	Flipflop status			
		$Q_3$	$Q_2$	$Q_1$	$Q_0$
		0	0	0	0
1	1	1	0	0	0
2	1	1	1	0	0
3	0	0	1	1	0
4	1	1	0	1	1

Upon the termination of  $CP_3$ , the contents of  $FF_3$  and  $FF_2$  are shifted to  $FF_1$  and  $FF_0$ , respectively; now,  $Q_3 = 0$ ,  $Q_2 = Q_1 = 1$ , and  $Q_0 = 0$ . Finally, at the end of  $CP_4$ , the word 1011 is stored in the register with  $Q_3 = 1$ ,  $Q_2 = 0$ , and  $Q_1 = Q_0 = 1$ .

To read out the word, four clock pulses are applied to the register. At the end of the fourth clock pulse,  $Q_3 = Q_2 = Q_1 = Q_0 = 0$ . Because the stored word is destroyed in reading out the shift register, the process is called *destructive readout* (DRO).



**Fig. 15.36** A four-bit serial-in, parallel-out shift register. (See Example 15.7.)

**example 15.7** Show how the serial shift register of Fig. 15.35 may be converted to a serial-in, parallel-out shift register (serial-to-parallel converter).

**solution** The circuit of a serial-in, parallel-out shift register is given in Fig. 15.36. Four AND gates are connected to the circuit. One input to each AND gate is connected to the output of a flipflop. The second input is connected to an *enable* signal. When the enable signal equals a logic 1, the content of each flipflop in the register is read simultaneously. Because the word in the shift register is not destroyed in reading out, the process is called *nondestructive readout* (NDR).

## 15.9 SEMICONDUCTOR MEMORIES

One important building block found in digital computers and systems is the memory. A memory is required for the storage of instructions (computer program), data, and results obtained in the processing of data. Whereas in the past ferrite cores were widely

used, semiconductor memories have replaced cores in most digital systems. In a semiconductor memory, flipflops are used to store a 0 or a 1 bit. The transistors in the flipflop may be bipolar (BIT) or field-effect (MOSFET).

Memories are organized to store  $W$  words, each word being  $B$  bits in length. The storage capacity of a memory, therefore, is  $W \times B$  bits. A typical capacity of a semiconductor memory chip is 4096 bits (1024 words  $\times$  4 bits/word). A group of bits is often referred to as a *byte*. Generally a byte is taken to be equal to 8 bits. Most semiconductor memories exhibit nondestructive readout.

### TYPES OF MEMORIES

Semiconductor memories may be categorized as:

1. Read-only memory (ROM).
2. Random-access memory (RAM).

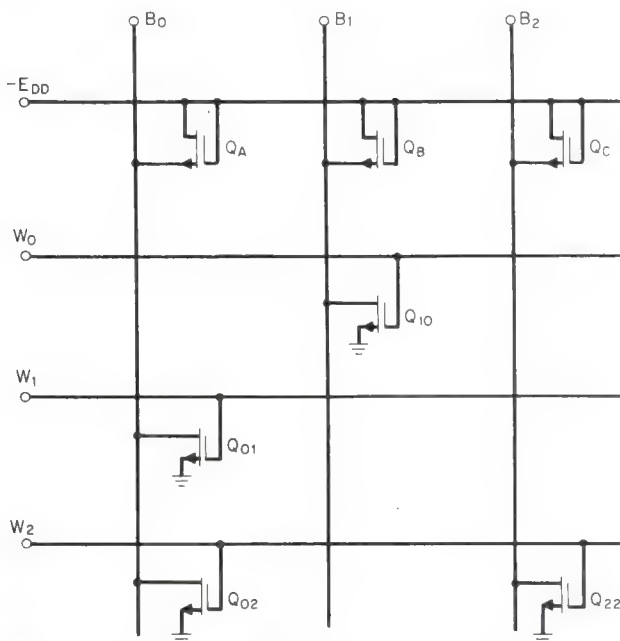
In the ROM, the data are written in once and thereafter cannot be changed. This type of memory, for example, may be used to store a computer program. Information stored in the RAM can be readily changed in the processing of data.

A variation of the ROM is the *programmable* ROM (PROM). In the PROM, each memory element is in series with an aluminum or Nichrome strip that acts as a fuse link. The user can open the links by passing a specified current through the element. In this manner elements may be removed to represent a 0 or a 1 bit.

Memories are also categorized as being either *static* or *dynamic*. A static memory requires no clock pulses; a dynamic memory needs clock pulses. Generally, dynamic memories are less costly and consume less power than static memories.

**ACCESS AND CYCLE TIMES** Two important terms used in characterizing memories are *access* and *cycle time*. Access time is the time needed to read out a word from the memory. Cycle time is equal to one over the rate at which a word may be selected for reading or writing. Typical access and cycle times are 400 and 600 ns, respectively.

**ROM** An example of a ROM using p-channel depletion-type MOSFETs is illustrated in Fig. 15.37. For simplicity, only three-bit ( $B_0, B_1, B_2$ ) and three-word ( $W_0, W_1,$



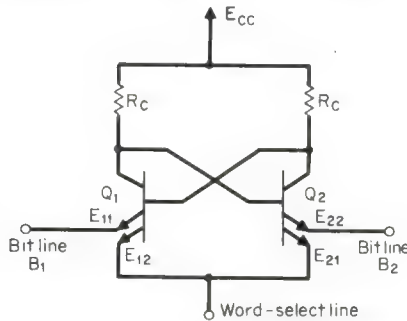
**Fig. 15.37** An example of a ROM employing p-channel depletion-type MOSFETs. For simplicity, only three bit and word lines are shown.

$W_2$ ) lines are shown. Typically, a semiconductor ROM has 1024 word lines, each word being four bits in length.

Transistors  $Q_A$ ,  $Q_B$ , and  $Q_C$  have their drains connected to the gate terminals. These transistors serve as the load resistors for the memory transistors. For example,  $Q_A$  is the load transistor for  $Q_{01}$  and  $Q_{02}$ ;  $Q_B$  is the load resistor for  $Q_{10}$ , and  $Q_C$  is the load resistor for  $Q_{22}$ . Each drain of memory transistors  $Q_{01}$ ,  $Q_{12}$ , etc., is connected to a bit line; their gates are connected to a word line. The presence of a transistor, such as  $Q_{10}$ , indicates a stored bit of data. An absence of a transistor indicates no stored bit of data.

For example, if word  $W_2$  is to be read, a negative voltage is applied to word line  $W_2$ . Because the transistors are p-channel devices,  $Q_{02}$  and  $Q_{22}$  are turned on. Bit lines  $B_0$  and  $B_2$ , therefore, read approximately 0 volts. Where a transistor is absent, a negative voltage is read. If a conducting transistor represents a 1 and an absent transistor a 0, the binary number read is 101.

**RAM** An example of a RAM cell using bipolar junction transistors is illustrated in Fig. 15.38. (Random-access memories using MOSFETs are also available.) Transistors  $Q_1$  and  $Q_2$  have multiple emitters, similar to that used in T<sup>2</sup>L. The circuit is a bi-stable MV in which the collector of one transistor is directly coupled to the base of the other transistor.



**Fig. 15.38** A typical memory cell using bipolar junction transistors found in random-access memories. In a four-bit, 1024-word RAM, for example, 4096 of these cells are used in the memory.

Bit lines  $B_1$  and  $B_2$  are connected to an emitter of each transistor. The word-select line is connected to the remaining emitters. Assume that when  $Q_1$  is conducting ( $Q_2$  is therefore OFF), a binary 1 is stored in the cell. If  $Q_2$  is ON ( $Q_1$  is now OFF), a binary 0 is stored.

In its quiescent state, the word-select line is returned to ground or a low voltage. If a binary 1 is stored in the memory cell, transistor  $Q_1$  is ON and  $Q_2$  is OFF. Current, therefore, flows through emitter  $E_{11}$ .

To read the contents of the cell an appropriate positive voltage is applied to the word-select line. Emitter  $E_{12}$  becomes reverse-biased with respect to the base of  $Q_1$ . Emitter  $E_{11}$ , however, is forward-biased owing to a small potential present at bit line  $B_1$ . As a result, emitter current flows in bit line  $B_1$ , indicating that a binary 1 was stored in the cell. Because  $Q_2$  is OFF, no current flows in bit line  $B_2$  when it is sensed.

To write or store data in the cell, the word-select line is raised to a suitable positive potential. If a small voltage is applied to bit line  $B_2$ , transistor  $Q_2$  is turned ON and  $Q_1$  is turned OFF. A binary 0 is thereby stored in the cell. If a large positive voltage is applied to bit line  $B_2$ , the reverse is true:  $Q_1$  is turned ON and  $Q_2$  is turned OFF. This corresponds to a binary 1.

## 15.10 DIGITAL-TO-ANALOG AND ANALOG-TO-DIGITAL CONVERTERS

In data processing systems, there is the need to convert a digital signal to its analog equivalent or an analog signal to its digital equivalent. To cite an example, consider

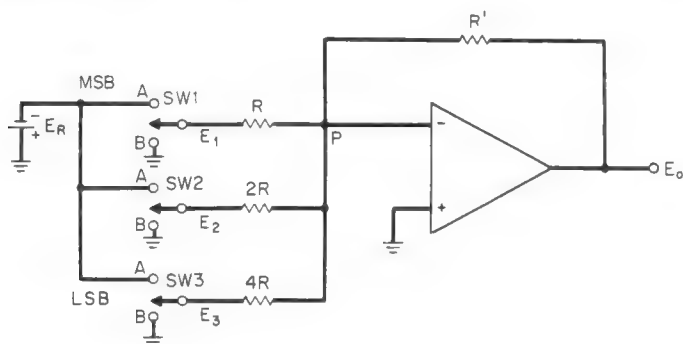


the digital voltmeter (DVM). To read a voltage, which is an analog quantity, the voltage is converted to a digital quantity by an analog-to-digital (A/D) converter. If the digital voltage is to be converted to an analog voltage, a digital-to-analog (D/A) converter is required. In this section, both types of converters are examined.

**D/A CONVERTER** An example of a D/A converter that converts a three-bit binary quantity to an equivalent analog quantity is illustrated in Fig. 15.39. The basic circuit of the D/A converter is a summing amplifier using an op amp (see Chap. 13). Output voltage  $E_o$  is equal to

$$E_o = \frac{-R'}{R} E_1 - \frac{R'}{2R} E_2 - \frac{R'}{4R} E_3 \quad (15.1)$$

For simplicity, mechanical single-pole, double-throw (SPDT) switches are shown in Fig. 15.39. When  $SW_1$  is in position A, it reads the most significant bit (MSB). Switch  $SW_3$  in position A reads the least significant bit (LSB). Any switch in position B is returned to ground.



**Fig. 15.39** An example of a weighted-resistor D/A converter. For simplicity, mechanical switches are shown.

Assume that the binary quantity  $N$  to be converted is  $N = 001$ . Let the feedback resistor  $R'$  be equal to  $R/2.5$  and  $E_R = -10$  V. For  $N = 001$ , switches 1 and 2 are returned to ground and switch 3 to  $E_R = -10$  V. By Eq. (15.1),

$$E_{o(1)} = \frac{-R/2.5}{4R} (-10) = 1 \text{ V}$$

Suppose that  $N = 011$  (decimal 3). Then, switches 2 and 3 are connected to  $E_R = -10$  V, and  $SW_1$  is returned to ground. By Eq. (15.1),

$$E_{o(3)} = \frac{-R/2.5}{2R} (-10) - \frac{-R/2.5}{4R} (-10) = 3 \text{ V}$$

Consider  $N = 111$  (decimal 7). All three switches are now connected to  $E_R = -10$  V, and

$$E_{o(7)} = \frac{-R/2.5}{R} (-10) - \frac{R/2.5}{2R} (-10) - \frac{R/2.5}{4R} (-10) = 7 \text{ V}$$

The preceding results indicate that the output voltage always corresponded to the binary input, and, indeed, a digital quantity was converted to an analog equivalent. For a four-bit quantity, another resistor equal to  $8R$  in series with a switch is connected to point P. In general, for an  $n$ -bit quantity,  $n$  resistors are required. The resistance value for the least significant bit is equal to  $2^{n-1} R$ .

The D/A converter of Fig. 15.39 is referred to as a *weighted-resistor converter*, because its output voltage depends on the values of the input summing resistors. These resistances have different values ( $R$ ,  $2R$ , and so on) and the resistor for the LSB can become excessively large ( $2^{n-1} R$ ).

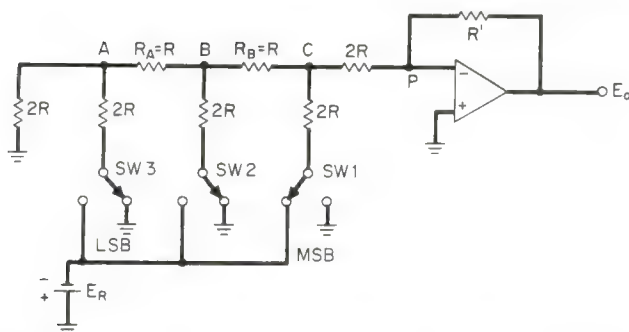


Fig. 15.40 A ladder-type  $D/A$  converter for a three-bit word. The values of resistors in this kind of converter are either  $R$  or  $2R$ .

Owing to these facts, it is difficult for all the resistors to track with temperature. If good tracking is not realized, errors result when the converter is operated at different temperatures. To overcome these difficulties, the *ladder-type*  $D/A$  converter of Fig. 15.40 has been developed. Although it uses twice as many resistors as the weighted-resistor type  $D/A$ , the values of the resistors are only either  $R$  or  $2R$ .

Assume that  $N = 100$  (decimal 4). Switch  $SW_1$  is connected to reference voltage  $E_R$ , and switches 2 and 3 are returned to ground. The resistance at node  $A$  with respect to ground is  $2R \parallel 2R = R$ . Addition of this resistance to  $R_A = R$  yields  $2R$ . The resistance at node  $B$  with respect to ground is therefore  $2R \parallel 2R = R$ . Addition of this resistance to  $R_B = R$  yields  $2R$ . Hence, looking to the left from node  $C$ , there is an equivalent resistance of  $2R$  to ground, as shown in Fig. 15.41a. Because the input voltage to the op amp. is approximately zero, point  $C$  also "sees" an equivalent resistance of  $2R$  to its right.

Current  $I$  flowing in the  $2R$  resistor connected to  $E_R$  is, therefore,

$$I = \frac{10}{2R + 2R \parallel 2R} = \frac{10}{3R}$$

The voltage at node  $C$  is current  $I$  multiplied by  $2R \parallel 2R$ :

$$\frac{10}{3R} \times 2R \parallel 2R = \frac{10}{3R} \times R = \frac{10}{3} \text{ V}$$

Using these results and letting  $R' = 2.4R$ , the model of Fig. 15.41b is drawn. By Eq. (15.1),

$$E_o = \frac{10}{3} \times \frac{2.4R}{2R} = 4 \text{ V}$$

which is the decimal equivalent of binary 100.

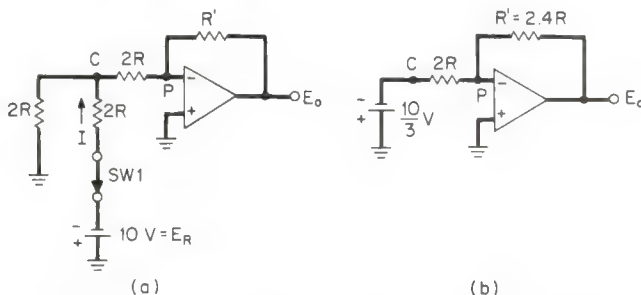
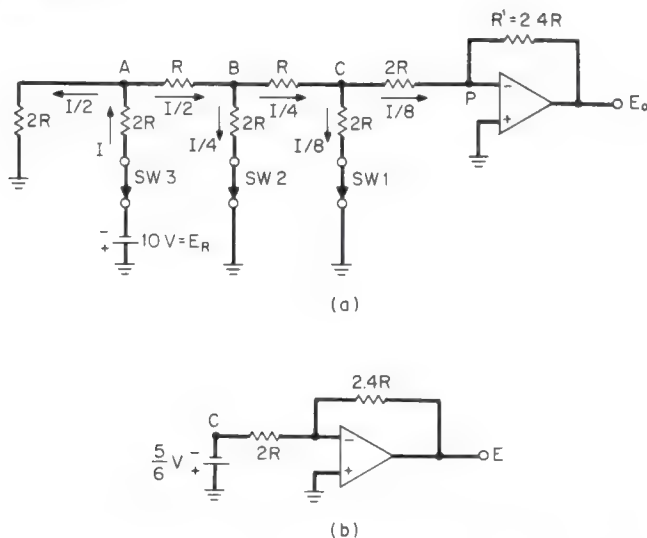


Fig. 15.41 A ladder-type  $D/A$  converter processing the binary number  $N = 100$ : (a) Resistance to the left or right of node  $C$  is equal to  $2R$ . (b) Model for determining output voltage,  $E_o$ .



**Fig. 15.42** Ladder-type  $D/A$  converter processing  $N = 001$ : (a) Circuit. (b) Model for determining  $E_o$ . (See Example 15.8.)

**example 15.8** In the ladder-type  $D/A$  converter of Fig. 15.40, determine  $E_o$  when  $N = 001$ . Assume that  $R' = 2.4R$ .

**solution** For  $N = 001$ , switches 1 and 2 are returned to ground and  $SW_3$  is connected to  $E_R$ . Referring to Fig. 15.42a, node A “sees” a resistance to its left of  $2R$  and an equivalent resistance of  $2R$  to its right. Because  $2R \parallel 2R = R$ , the current in  $SW_3$  is  $10/(2R + R) = 10/3R$  amperes.

At node A, half of current  $I$  ( $I/2$ ) flows to the left, and the other half flows to the right. Reaching point B, the current is divided again by 2, and  $I/4$  flows toward node C. At C it is split again, and  $I/8$  flows through  $2R$  in series with  $SW_1$ .

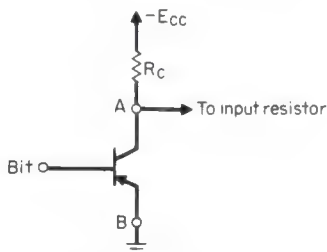
The voltage at node C with respect to ground is equal to the product of  $I/8$  and  $2R$

$$\frac{I}{8} \times 2R = \frac{1}{8} \times \frac{(-10)}{3R} \times 2R = \frac{-5}{6} \text{ V}$$

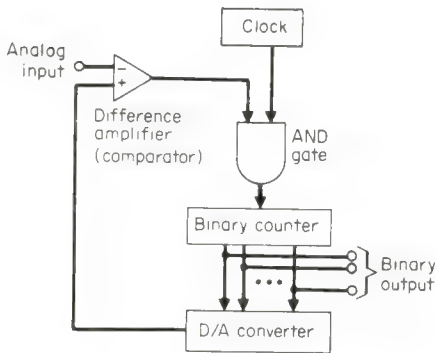
This is illustrated in the model of Fig. 15.42b. By Eq. (15.1)

$$E_o = \left( \frac{-2.4R}{2R} \right) \left( \frac{-5}{6} \right) = 1 \text{ V}$$

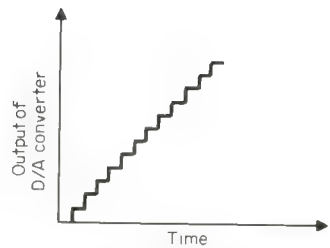
Electronic switches are used in place of mechanical switches. To gain some idea how this is accomplished, consider the transistor switch of Fig. 15.43. As in the mechanical switch, point A goes to an input resistor and point B to ground.



**Fig. 15.43** An elementary pnp transistor switch used in place of mechanical switches in  $D/A$  converters.



**Fig. 15.44** An example of an A/D converter. The digital (binary) output is obtained from the binary counter.



**Fig. 15.45** A ramp (staircase) waveform at the output of the D/A converter of Fig. 15.44.

When a bit is present at the Bit terminal, assume that a positive voltage is impressed across the base of the transistor. Because the transistor is pnp, it is cut off, and the input resistor is connected to reference voltage  $-E_{CC}$  through resistance  $R_C$ . If a bit is not present, a negative is impressed across the base. The transistor is turned ON, and the input resistor is returned to ground.

Because of collector resistor  $R_C$ , some error is introduced in the converter. Other configurations, containing a pair of bipolar or field-effect transistors, are used for minimizing the error.

**A/D CONVERTER** One example of a commonly used A/D converter is provided in Fig. 15.44. One input to the AND gate is derived from the clock, and the other input is obtained from the output of a difference amplifier. The difference amplifier compares the output of the D/A converter and the analog input signal. For this type of operation the difference amplifier is referred to as a *comparator*.

If the output of the D/A converter is less than the analog signal, the AND gate remains operative, and the binary counter continues counting clock pulses. When their difference is zero, the AND gate becomes inoperative, and the binary counter stops counting. The output of the counter is a binary number equal to the analog input signal.

The output of the D/A converter continuously increases by one bit as the binary counter counts. The resultant waveform, referred to as a *ramp*, or *staircase*, waveform is illustrated in Fig. 15.45. For this reason, the circuit of Fig. 15.44 is referred to as a ramp A/D converter. Although other methods are used in analog-to-digital conversion, the ramp type is fairly common.

## 15.11 THE MICROPROCESSOR

A microprocessor may be regarded as a general-purpose digital computer that is available on a few silicon chips. Currently, over a dozen IC manufacturers are producing more than two dozen different types of microprocessors. Although most microprocessors use MOSFETs, a few employ the BJT. Applications of the microprocessor include traffic control systems, numerically controlled machine tools, and the automatic testing of engines.

For illustrative purposes, the microprocessor chips manufactured by National Semiconductor Corporation are considered. National produces two basic microprocessor chips. These are the Register, Arithmetic, and Logic Unit (RALU); and the Control Read-Only Memory (CROM).

The RALU performs the basic control and arithmetic functions in processing data. The CROM contains the instructions (program) that direct the execution of data by the RALU. In addition, National provides software, that is, programs for a given application of the microprocessor. By using a number of RALU and CROM chips, it is possible to realize a 16-bit word computer, referred to as a *microcomputer*.



# Chapter 16

## Power Supplies

### 16.1 INTRODUCTION

The function of an electronic power supply is to convert the available prime power source into a form required by a particular system. Typical prime power sources are rated at: 115 V, 60 Hz, single phase; 208 V, 60 Hz, three phase; 115 V, 400 Hz, single phase; and 12 or 28 V, dc.

The voltage requirements will vary depending on the type of system and also the portion of the system. Solid-state systems require considerably lower dc voltages than vacuum-tube systems. Some systems require regulated dc voltages, while others do not. In a television receiver, the r-f and video portions may require low to moderate values of dc voltages, while the picture tube may require high voltages in the order of 30 000 V. In computers, low dc voltages may be required for solid-state logic circuits, while intermediate values of dc voltages may be required for solid-state analog circuits.

Regardless of the type of electronic power supply involved, the basis of the supply is its rectifier system. We will begin our study of power supplies with a review of the basic types of rectifier systems.

### 16.2 RECTIFIER SYSTEMS

A block diagram of an electronic power supply is shown in Fig. 16.1. The ac power source is generally the power delivered by the power company, but in some applications other sources may be used. One example is in automotive electric systems that employ alternators. An alternator is basically an ac power source, and the ac power is converted to dc power by diodes which are mounted on the metal frame of the alternator. In this particular example the diodes are semiconductor types, but in other applications vacuum-tube diodes may be used.

Returning to the block diagram, the ac power is delivered to the primary of a power transformer. This transformer may be a step-up or a step-down type, depending on how much voltage is required by the load. Some electronic power-supply circuits operate without a transformer, and are called *transformerless power supplies*. Such supplies are less expensive to build, and this is considered to be an important advantage. However, the amount of voltage delivered to the load is a function of the voltage at the power source, and therefore, the use of such transformerless supplies is limited.

The output of the power-supply transformer is delivered to a section marked "diode rectifier circuit." The purpose of the rectifiers is to convert the ac voltage input to a pulsating dc voltage. Diodes function as rectifiers because they will allow current flow in only one direction—that is, from cathode to plate or anode. When an ac voltage is placed across the diode, it conducts during the half-cycles when the plate (or anode) is



## 16-2 Power Supplies

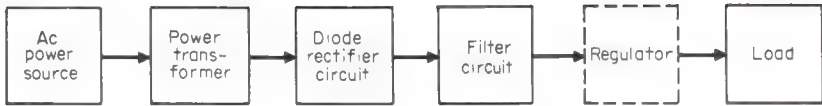


Fig. 16.1 Block diagram of an electronic power supply.

positive with respect to the cathode, but it will not conduct on the alternate half-cycles when the plate is negative with respect to the cathode.

The output of the diode rectifier circuit is a pulsating voltage. It cannot be used for operating transistor and vacuum-tube circuits because these circuits normally require an unvarying dc voltage. Therefore, a filter circuit is used to smooth out the voltage ripples. Usually filter circuits are *passive*, and are comprised of resistors and capacitors, or inductors and capacitors. (A passive filter is one that simply operates as a smoothing circuit.) There is always some loss of power in such a circuit, but this is not considered to be a serious disadvantage since the input stages of the power supply can be designed to deliver a sufficient amount of power for the filter loss, as well as for the load.

In the block diagram there is a dotted section marked *regulator* between the filter circuit and the load. Not all power supplies have regulators. Their purpose is to maintain a constant value of voltage across the load regardless of changes in the load resistance or ac input voltage. The regulator is usually an electronic circuit that accomplishes its function by sensing any change in the voltage across the load, and then feeding this change into the regulator circuit for use as a control to offset the change.

The last stage in the block diagram of Fig. 16.1 is the *load*. The load is the component or circuit that receives the dc power. In a radio or television set the load will be the vacuum tubes or transistors operating in the electronic circuitry. In other applications the load may be a dc motor, a battery that is being charged, or any other circuit or component requiring a dc source of power.

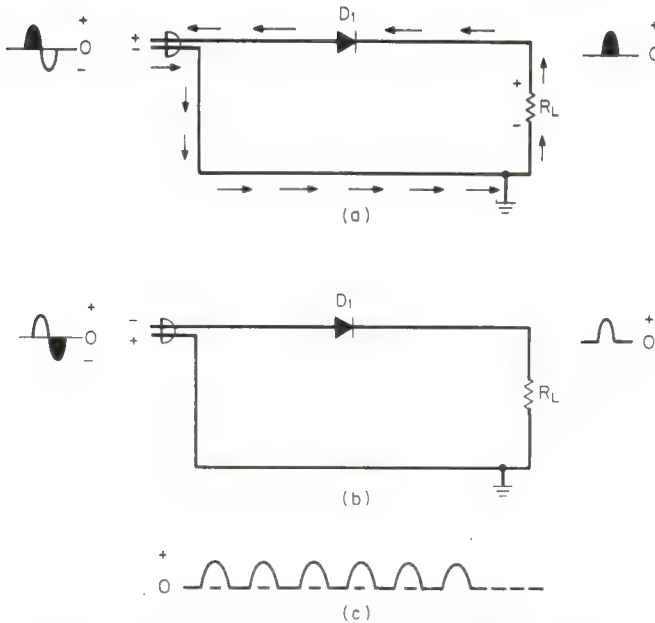
From the discussion of the block diagram, it can be seen that the term “electronic power supply” is somewhat misleading. This circuit does not *supply* the power, but rather, converts it from one form (ac) to another form (dc). However, the name is widely used in electronics. Transformers and filter circuits are covered in other chapters in this book. Our concern here will be primarily with various rectifier configurations, how the rectifier circuits operate, and the advantages and disadvantages of each circuit.

**HALF-WAVE RECTIFIERS** The simplest rectifier circuit is the one-diode configuration shown in Fig. 16.2. This is a half-wave rectifier, so named because the output voltage waveform for the circuit is only one-half of the ac input voltage waveform. It is an example of a transformerless power supply, and it is very popular in low-cost home-entertainment electronic systems such as table-model radios and record players.

The input voltage to the circuit is an ac waveform. On one-half cycle, which we have designated the *positive half-cycle*, the anode of the diode is made positive and electron current can flow through the diode. The path of electron current flow is shown by the arrows in Fig. 16.2a. It will be noted that the current is flowing upward through the load resistor  $R_L$ . This means that the voltage drop across  $R_L$  will be such that the negative polarity of the voltage is at the bottom and the positive polarity is at the top. The output waveform during the positive half-cycle is an exact reproduction of the positive half-cycle of input waveform.

During the next half-cycle, the input voltage is negative. (See Fig. 16.2b.) This causes a negative voltage to be applied to the anode of  $D_1$ . No electron current can flow through the diode during this half-cycle, and the output voltage will be zero, as shown by the waveform on the diagram. Figure 16.2c shows the waveform of the voltage across the load resistance.

The waveform of Fig. 16.2c is referred to as a *pulsating dc voltage*. It is a dc voltage because its polarity is always positive (or zero). It is pulsating because the amplitude changes from moment to moment. Such a waveform is not useful for operating tube or transistor circuits unless it is first filtered. The purpose of filtering is to take out the pulsations. Because the voltage rises from zero to its maximum value and back to zero, and then stays at zero for a certain period of time, makes this waveform difficult to filter.



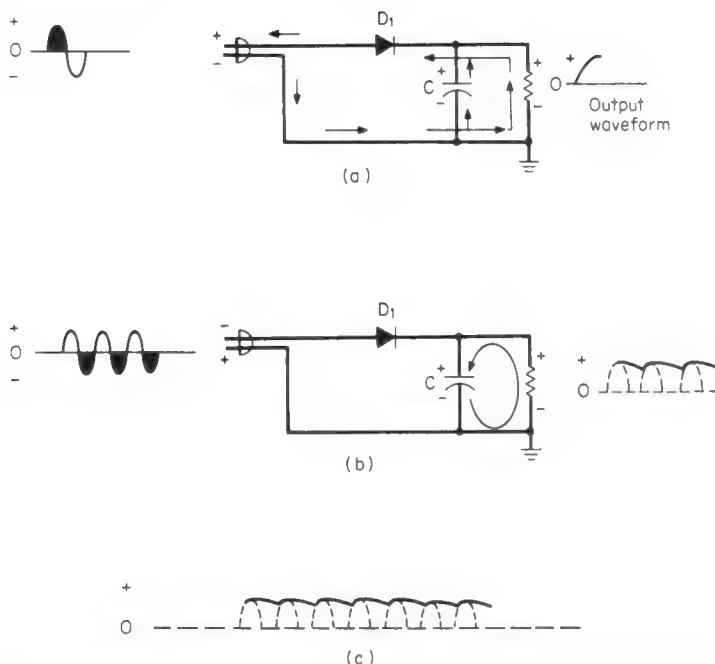
**Fig. 16.2** The half-wave rectifier circuit with current flow and voltage waveforms: (a) Current flows during the positive half-cycle. (b) No current flows during the negative half-cycle. (c) Output voltage waveform of a half-wave rectifier circuit.

If this waveform is going to be converted to a steady dc voltage, then all the periods during which the voltage is zero must be filled in.

**FILTERING** The procedure for filtering the voltage usually involves charging a capacitor to the *peak* ac input voltage during a time when the diode is conducting and then having the capacitor discharge through the circuit in order to maintain the voltage across the load. To understand this, we will look at a simple filter circuit consisting only of a capacitor, as shown in Fig. 16.3. The only difference between this half-wave rectifier circuit and the one shown in the previous illustration is that a filter capacitor  $C$  has been added. This filter capacitor is an electrolytic type with a high capacitance value which enables it to store a considerable amount of energy during each half-cycle. The positive sign beside the capacitor symbol indicates that the capacitor is of the electrolytic type.

During the positive half-cycle of input voltage, the diode conducts as before. In this case, however, only a small part of the conducting current initially flows through the load, producing the output voltage waveform. The rest of the current flows into one plate of the capacitor and out of the other, thus charging the capacitor. This is shown in Fig. 16.3a. During the negative half-cycle, as shown in Fig. 16.3b, a negative input voltage is delivered to the plate of the diode, and it can no longer conduct. At this time the voltage across the capacitor causes current to flow through the load resistance. The capacitor discharge current is indicated by the arrow in Fig. 16.3b. This current prevents the voltage from dropping to zero, as it normally would without the presence of the capacitor.

The filter capacitor charges to the peak value of the ac input voltage. When the value of input voltage begins to decrease below the voltage across the capacitor, then the capacitor begins to discharge through the resistor. Figure 16.3c shows the filtered output waveform from the power supply in solid lines. The unfiltered output waveform (without the filter capacitor) is shown with dotted lines. Although the voltage drops

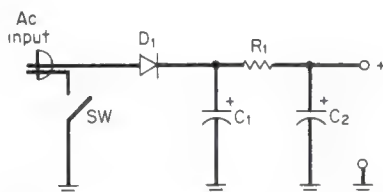


**Fig. 16.3** The half-wave rectifier with a simple filter: (a) During the first positive half-cycle the capacitor charges. (b) During the next half-cycle the capacitor discharges. (c) The filtered output waveform.

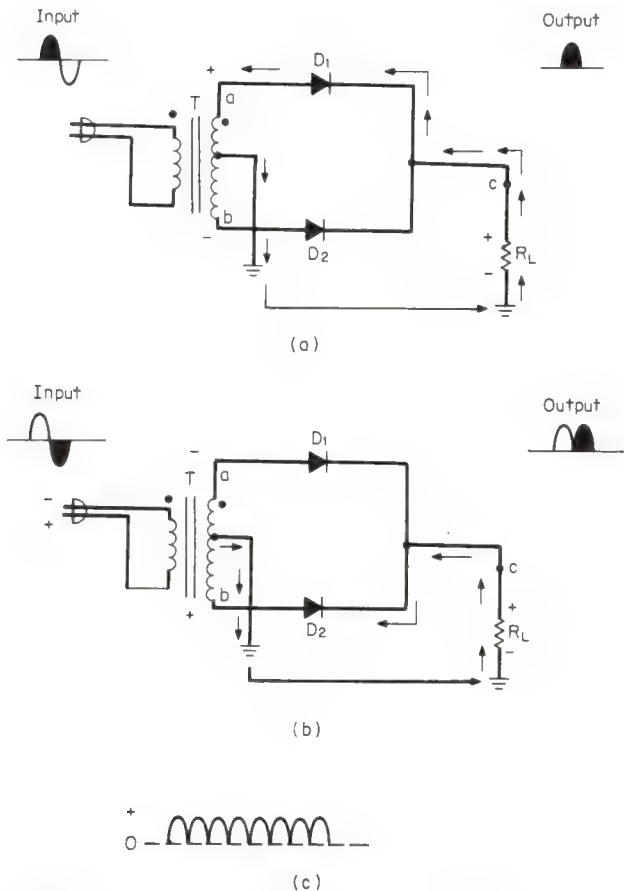
somewhat during the negative half-cycles of input voltage, it does not drop all of the way to zero.

The operation of the simple capacitor filter shown in Fig. 16.3 is based on the fact that a capacitor can store energy. Basic theory tells us that a capacitor stores energy in the form of an *electrostatic field*, and an inductor (another device that can store energy) stores it in the form of an *electromagnetic field*. Passive filter circuits always rely on the ability of either a capacitor or an inductor (or both) to store energy and return the energy to the circuit at the proper time. Figure 16.4 shows a half-wave rectifier with a more elaborate filter. This circuit is popular with low-cost electronic circuits. The switch (SW) energizes or deenergizes the circuit. The pi filter, comprised of  $R_1$ ,  $C_1$ , and  $C_2$ , smooths the dc pulsations from the half-wave rectifier. This circuit is used mainly for low-current applications, unless the capacitors are very large in value. In the latter case, the resistor can be kept to a reasonably low value, to minimize the voltage drop across it.

**FULL-WAVE RECTIFIERS** One of the most important disadvantages of the half-wave rectifier circuit is the difficulty in filtering its output waveform. The relatively



**Fig. 16.4** Half-wave rectifier with a filter.



**Fig. 16.5** The full-wave rectifier: (a) Full-wave rectifier showing current flow during one half-cycle. (b) Full-wave rectifier showing current flow during the next half-cycle. (c) Output waveform of the full-wave rectifier.

long period between positive voltage peaks makes it necessary for the filter circuit to supply a considerable amount of current flow through the load for a long period of time—that is, the discharge time is long compared to the charge time of the filter components. In spite of this disadvantage, the half-wave rectifier is a popular circuit because it is inexpensive and relatively trouble-free.

Figure 16.5 shows a full-wave rectifier circuit. This type of circuit *requires* a transformer  $T_1$  with a center-tapped secondary winding. In addition to the need for the transformer, the circuit also requires the use of two diodes—shown as  $D_1$  and  $D_2$  in the circuit.

With an ac voltage across the primary of the transformer, the polarity of the voltage across the secondary will periodically reverse. The dot notation on the transformer is a standard way of indicating points of identical phase. When the voltage on the primary winding goes positive at the point where the dot is located, then at that instant the voltage at the secondary also goes positive at the point where the dot is located.

On one half-cycle, point *a* will be positive; and on the next half-cycle, point *a* will be negative. The voltage at point *b* will always have the opposite polarity of the voltage at point *a*. The center tap which is connected to the common, or *ground point*, will always be maintained at zero volts.

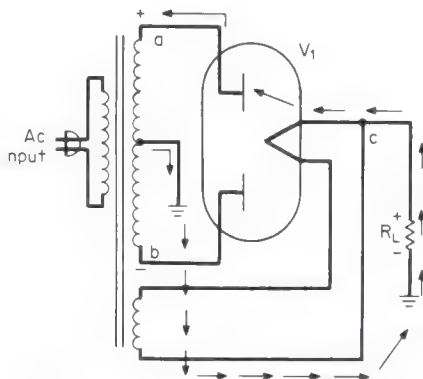
Figure 16.5a shows the operating condition during the half-cycle when point  $a$  is positive and point  $b$  is negative. The negative voltage on the plate of  $D_2$  prevents it from conducting. The positive voltage on the plate of  $D_1$ , however, allows that tube to conduct. The conduction path is shown by arrows. If point  $a$  is positive, and if the center tap is zero, then it stands to reason that the center tap is negative with respect to point  $a$ . Starting at the negative terminal of the voltage source (which in this case is one-half the secondary winding) the electron current flows out of the center tap into the common circuit. It then flows out of the common circuit, through  $R_L$ , and to the junction of the two cathodes. The electron current cannot flow through  $D_2$  because it is cut off with a negative voltage on its anode. Therefore, all the current flows through diode  $D_1$ . This completes the circuit.

During the first half-cycle, the current is seen to be flowing upward through the load resistor, making the voltage drop across it positive at point  $c$ . The output waveform shown on the illustration is seen to be a reproduction of the input waveform during this half-cycle. The amplitude of the output may be either smaller or larger than the input, depending on the turns ratio of the power transformer. If the turns ratio is 1:1, it means that the number of turns on the primary and on *one-half* of the secondary are equal. In such a case, and assuming there is no filter circuit, the output amplitude will be approximately equal to the amplitude to the input signal. It is necessary to say "approximately" because there is always some voltage drop across the diode itself. The actual value of this drop can be obtained by multiplying the current through the diode by the forward resistance of the diode. Although this voltage drop may be small enough to be neglected in most cases, it does, nevertheless, reduce the output amplitude to something less than the input voltage (measured across one-half of the transformer secondary winding). There is also a voltage drop across the diode in the half-wave rectifier circuit previously discussed.

In Fig. 16.5b the circuit is shown with the voltage polarity across the secondary reversed. In this case, the anode of  $D_1$  is negative (with respect to ground) and the anode of  $D_2$  is positive. Therefore, only  $D_2$  can conduct during this half-cycle. Again, electron current leaves the center tap, flows into the common connection, and then through the load resistor  $R_L$ . The current flow is shown by arrows. It is important to note that the electron current is again flowing in the same direction as before, through  $R_L$ , making point  $c$  again positive with respect to ground. This means that a positive-going voltage will appear across the load for *both* half-cycles of input voltage.

Figure 16.5c shows the output waveform of a full-wave rectifier. Since there are no long periods during which zero voltage is generated across the load, this waveform is easier to filter than the half-wave rectifier output. That is one of the important advantages of the full-wave rectifier circuit.

**TUBE RECTIFIERS** Some full-wave rectifier power supplies are designed with diode tubes having filament emitters. Filament emitters are more efficient, in general,



**Fig. 16.6** A full-wave rectifier circuit using tubes with filament emitters.



than cathode-type emitters. Cathodes are used whenever it is necessary to isolate the cathode emitter from the heater circuit. Figure 16.6 shows a full-wave rectifier with the filament emitters. The filament voltage is obtained from a separate winding on the transformer. This is a common practice with full-wave rectifier circuits. In other rectifier circuits, there may be a separate filament transformer, or the filament may be in series with a number of other filaments which produce (when in series) a total voltage drop equal to the line voltage. When the latter method is used, only tubes with cathode emitters can be used, so that the ac filament circuit is isolated from the dc supply voltage. (The only purpose of the filament winding in the circuit of Fig. 16.6 is to provide current for heating the filament. It does not enter into the operation of a rectifier power supply in any other way.)

The circuit shows the current flow for one-half cycle, during which point *a* is positive and point *b* is negative. As before, only the plate with the positive voltage can attract electrons. Therefore, only that half of the tube conducts during this half-cycle. Electron current leaves the center tap, flows upward through the load resistor into the filament, and then from the filament to the plate of the upper diode. During the next half-cycle the upper diode will have a negative voltage on its plate, and the lower diode will have the positive voltage. Therefore, the current flow will be through the lower diode.

An important consideration here is the fact that the filament winding of the transformer will be above ground by the amount of voltage drop across  $R_L$  because of the common connection at point *c*. In some power supplies this may be 300 or 400 V (or more) above ground. In choosing a transformer for this type of circuit, it is important to determine that the transformer secondary can withstand the resulting voltage stress. If the filament winding is above ground by 400 V, and point *b* is negative by 400 V during one-half cycle, then the voltage difference between the filament winding and that half of the winding is 800 V. Although this is not an excessively high voltage, it does, nevertheless, put a voltage stress between windings within the transformer itself and the transformer core, which is grounded.

Although the output waveform of the full-wave rectifier is easier to filter, this circuit is not without disadvantages. The fact that the secondary winding of the transformer must be *center-tapped* means that only one-half of the total secondary voltage will appear across the load. As an example, in the circuit of Fig. 16.6 assume that the voltage between point *a* and point *b* is 1000 V. This would be the total secondary winding voltage. Under this condition the voltage across  $R_L$  would be only 500 V. (Again, we are neglecting the drop across the diode itself.)

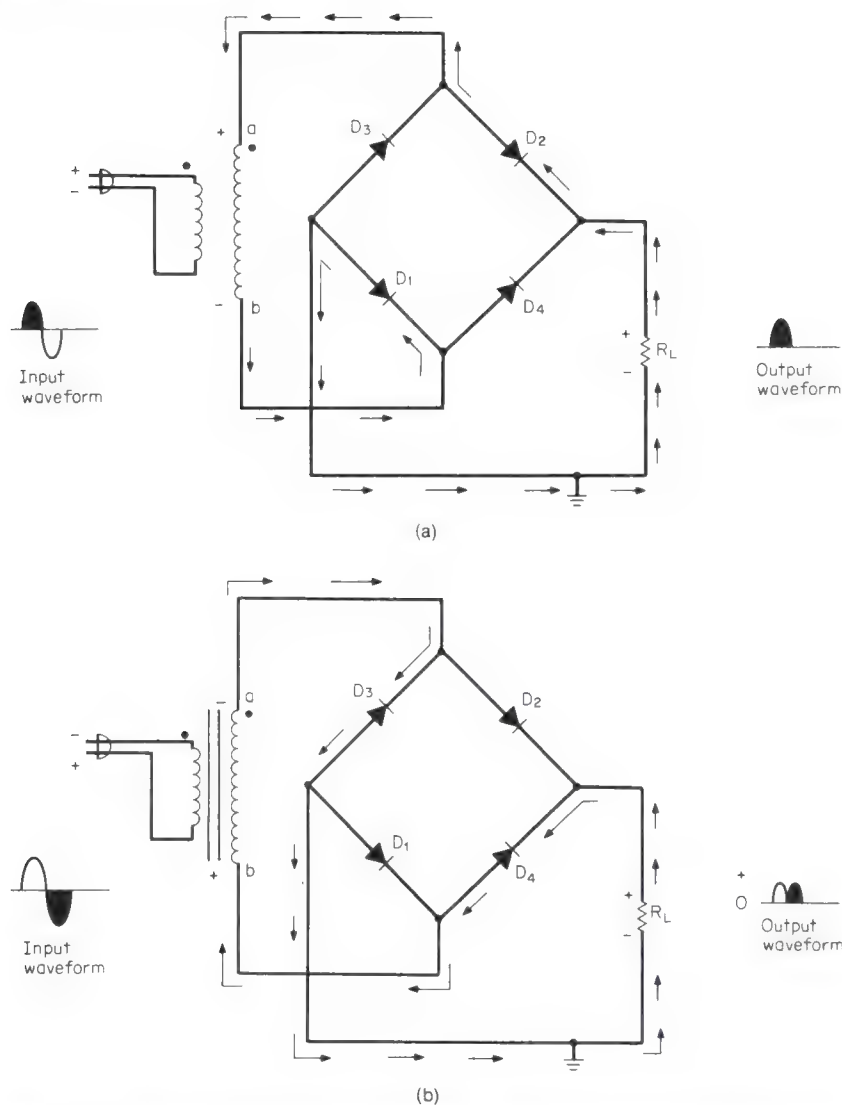
Another disadvantage that must be taken into consideration is that two diodes are needed for the operation of the full-wave rectifier. This means an added expense and somewhat reduced circuit reliability.

**BRIDGE RECTIFIERS** A full-wave rectifier circuit that makes use of the *full* transformer secondary-winding voltage, and does not require the center tap, is shown in Fig. 16.7. It is called a bridge rectifier because of its similarity to the configuration of the Wheatstone bridge. Although a transformer is shown in the circuit, it may also be connected as a transformerless power supply.

On the first half-cycle, shown in Fig. 16.7*a*, point *a* becomes positive with respect to point *b*. If we consider the secondary winding of the transformer as the voltage source, we can trace the electron current flow (starting at point *b*) for this half-cycle. The arrows show the current path flowing from *b* through  $D_1$ , up through  $R_L$ , and through  $D_2$  back to point *a* on the secondary winding. Diodes  $D_3$  and  $D_4$  are cut off during this half-cycle, because  $D_3$  has a positive voltage on its cathode, and  $D_4$  has a negative voltage on its anode. In order for a diode to conduct, it is necessary for the anode to be positive with respect to the cathode. If the cathode of a diode is made highly positive with respect to the anode, then this is the same as saying that the anode is negative with respect to the cathode. That is why the positive voltage on the cathode of  $D_3$  prevents it from conducting.

The next half-cycle of operation is shown in Fig. 16.7*b*. The polarity of voltage across the secondary winding has now been reversed, making point *a* negative with respect to point *b*. Starting at the negative terminal of the source (which is now point *a* on the secondary winding), current flows through  $D_3$ , through  $R_L$ , through  $D_4$ , and back to point *b* on the secondary winding. Diode  $D_1$  cannot conduct because of the positive





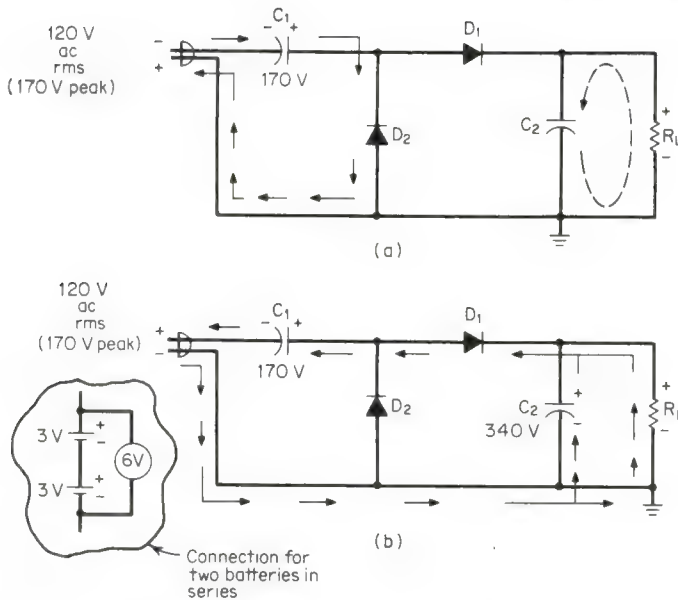
**Fig. 16.7** A bridge rectifier circuit: (a) Current flow in the bridge rectifier circuit for the first half-cycle of operation. (b) Current flow during the second half-cycle.

voltage on its cathode, and  $D_2$  cannot conduct because of the anode during this half-cycle.

The diodes in the bridge rectifier circuit may be thought of as switches which are operated by voltage polarities. During both half-cycles of operation, the current is seen to flow in the same direction through  $R_L$ . This means that the circuit produces a full-wave pattern like the one shown in Fig. 16.7b.

When semiconductor diodes are used for bridge rectifiers, they are often mounted in the same package for convenience.

The bridge rectifier circuit is found in many applications in the communications field and elsewhere. It is also used in some instrument circuits. Most meter move-



**Fig. 16.8** The half-wave voltage doubler circuit: (a) This half-cycle charges  $C_1$  to 170 V. (b) During this half-cycle, the 170 V across  $C_1$  is in series with the 170-V input, making a total of 340 V; capacitor  $C_2$  charges to 340 V.

ments on voltmeters respond only to a dc voltage. In order to measure an ac voltage, it is necessary to convert the alternating current into a direct current. A half-wave rectifier circuit *could* be used, but the bridge rectifier is generally employed because it is a full-wave rectifier. The input voltage—that is, the voltage being measured—is applied directly across the bridge at points *a* and *b*, and the meter movement is connected into the circuit at the position of  $R_L$  in the circuit of Fig. 16.7.

**HALF-WAVE VOLTAGE DOUBLERS** The half-wave voltage doubler circuit of Fig. 16.8 will provide an output voltage across  $R_L$  that is approximately twice the peak input ac voltage to the circuit. Thus we have a transformerless circuit which steps up the line voltage and converts it to a dc potential.

During one-half cycle of input power, the voltage polarity will be as shown in Fig. 16.8a. The positive voltage on the anode of  $D_2$  causes it to conduct, and the conduction path is shown by the arrows. This charges  $C_1$  to the peak voltage (170 V in this example). At the same time, diode  $D_1$  cannot conduct because its anode is not positive with respect to its cathode.

The next half-cycle of operation is illustrated in Fig. 16.8b. This shows that the voltage across capacitor  $C_1$  and the input voltage are in series. In other words, they are connected plus to minus as are the two batteries in series shown at the inset. The input voltage to the circuit during this half-cycle is equal to the voltage across  $C_1$  (170 V) plus the peak voltage across the line, making a total of 340 V. The negative side of the line is connected to the anode of  $D_2$  which prevents that diode from conducting during this half-cycle. The positive voltage on capacitor  $C_1$  is applied to the anode of  $D_1$ , and therefore  $D_1$  can conduct during this half-cycle. The conduction path is shown by the arrows.

It will be noted that the current in Fig. 16.8b divides: part of it flowing into capacitor  $C_2$ , and the remainder flowing through the load. Capacitor  $C_2$  will charge to the full 340 V during this half-cycle (minus, of course, the drop across the diode, which can usually be neglected in this type of circuit).

On the next half-cycle, the condition of Fig. 16.8a again prevails. The input voltage

recharges capacitor  $C_1$  by its conduction through  $D_2$ . At the same time  $D_1$  is cut off during this half-cycle. It will be remembered that capacitor  $C_2$  was charged to the 340-V input during the half-cycle shown in Fig. 16.8b. Now capacitor  $C_2$  will discharge through the load, as shown by the dotted arrow. This prevents the voltage across the load from dropping to zero, as it would without the filter capacitor.

The output voltage waveform is a filtered half-wave voltage like the one shown in Fig. 16.3c. If the value of load resistance in this circuit is small, then capacitor  $C_2$  will discharge rapidly through it, causing the voltage across  $R_L$  to go to zero during some portion of the half-cycle when  $D_1$  is not conducting. Therefore, this type of power supply cannot be used for circuits that require a high current flow—that is, circuits that load the supply. Both  $C_1$  and  $C_2$  are normally quite large in capacitance value, and are usually electrolytic capacitors. However, even with a large capacitance value for  $C_2$ , the power-supply circuit cannot sustain a heavy load current.

If  $R_L$  were a variable resistor, it would be found that the smaller the value of resistance, the lower the output voltage for the circuit. What is more important is that when the load current through  $R_L$  varies from moment to moment, the output voltage of the power supply also varies. An ideal power supply would be one that maintains a constant output voltage, regardless of the amount of load current being drawn. Such a power supply is not realizable in practice, but can be approached closely under certain conditions. The *regulation* of the power supply is a measure of how well it maintains its output voltage under variable load conditions. The disadvantage of the half-wave voltage doubler shown in Fig. 16.8 is that it has poor regulation. Therefore a half-wave voltage doubler is not generally used where the load varies widely.

**FULL-WAVE VOLTAGE DOUBLERS** The distinguishing feature of the half-wave voltage doubler just described is that it delivers power to the load only during one-half cycle of input alternating current. This, in part, accounts for the relatively poor regulation of the circuit. A full-wave doubler provides better regulation and is easier to filter.

A full-wave voltage doubler is shown in Fig. 16.9. On the first half-cycle, the input voltage will have a polarity as marked at the plug in Fig. 16.9a. A positive voltage, under this condition, is delivered to the cathode of  $D_1$ , and that diode is cut off. At the same time, the anode of  $D_2$  receives a positive voltage, and that diode *will* conduct. The conduction path is shown by the arrows. Capacitor  $C_1$  charges to the peak line voltage (170 V) during this period of time.

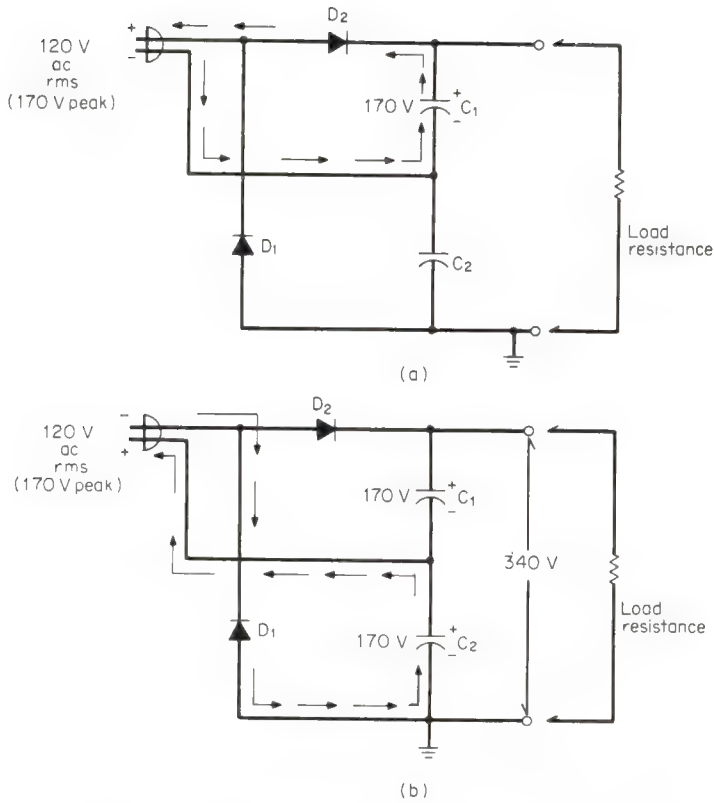
On the second half-cycle, the polarity of the voltage at the plug reverses—placing a negative voltage on the anode of  $D_2$ , and also a negative voltage on the cathode of  $D_1$ . (See Fig. 16.9b.) Under this condition, only  $D_1$  will conduct. The conduction path is again shown by arrows. Capacitor  $C_2$  charges to the peak line voltage of 170 V. The voltage across  $C_2$  is added in series to the voltage across  $C_1$  (which is obtained during the previous half-cycle). The two voltages, in series, combine to produce a 340-V output voltage. The load resistance is connected across the two series capacitors.

It might be argued that this is not really a full-wave rectifier system because the total voltage is not developed at the output capacitor circuit on *both* cycles of input. Instead, the voltage is delivered to one-half of the output circuit each time the input voltage reverses polarity. However, at least *some* voltage is delivered to the output capacitor circuit each half-cycle, and this makes it more of a full-wave rectifier than the one shown in Fig. 16.8.

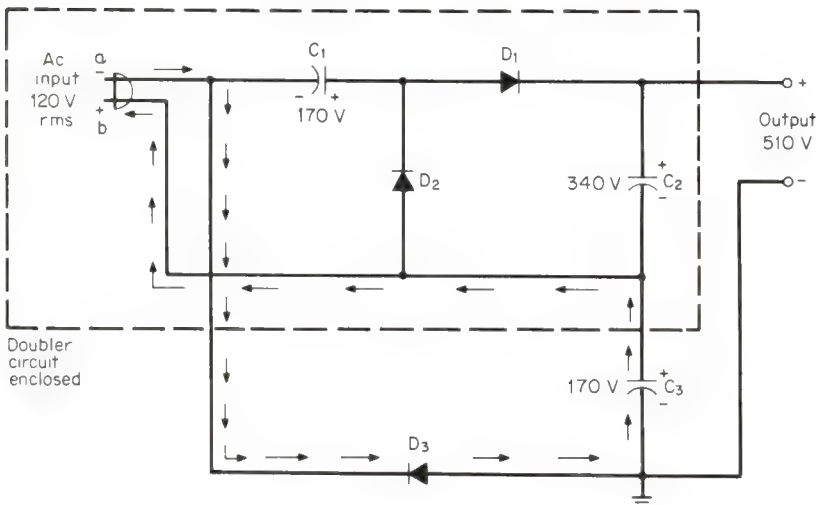
Although the regulation of the full-wave voltage doubler circuit is somewhat improved over that of the half-wave doubler, the circuit is still unable to sustain a heavy load current without a serious drop in the output voltage.

**VOLTAGE TRIPLERS AND OTHER VOLTAGE MULTIPLIERS** In the voltage doubler circuits just discussed, the principle of operation is based on charging a capacitor in one-half cycle, and then adding the voltage across that capacitor to the circuit voltage (or to the voltage across another capacitor) during the next half-cycle. It should be obvious that any number of capacitors can be charged, and their voltages added. However, as the amount of voltage multiplication goes up, the amount of regulation of the power-supply circuit becomes poorer and poorer.

Figure 16.10 shows another example of a voltage multiplier. (Voltage multipliers are diode rectifier circuits that increase the input voltage. The *half-wave doubler*, and



**Fig. 16.9** The full-wave voltage doubler: (a) First half-cycle of operation. (b) Operation during the second half-cycle.



**Fig. 16.10** A voltage tripler.

the *full-wave doubler* are examples. Other diode circuits that multiply the input voltage are triplers, quadruplers, quintuplers, etc.) The circuit of Fig. 16.10 is a voltage tripler. The portion of the circuit enclosed in dotted lines is exactly the same as for the half-wave voltage doubler previously discussed, and illustrated in Fig. 16.8. Capacitor  $C_2$  is charged to twice the peak input line voltage in this part of the circuit. Diode  $D_3$  and capacitor  $C_3$  have been added to the doubler circuit to convert it into a tripler.

On one-half cycle, point  $a$  on the plug will be negative with respect to point  $b$ , and this places a negative voltage on the cathode of  $D_3$ . The electron current flow through  $D_3$  starts at point  $a$ , flows through the diode up into capacitor  $C_3$ , and then out of the capacitor and back to point  $b$ . This electron current path is indicated by solid arrows on Fig. 16.10. During the same half-cycle when  $C_3$  is charged,  $D_2$  is also conducting and charging capacitor  $C_1$ .

On the next half-cycle point,  $a$  is positive with respect to point  $b$ , and the diode  $D_3$  is cut off. Diode  $D_1$  conducts, and charges capacitor  $C_2$  to twice the peak line voltage, as discussed previously for the doubler circuit.

The overall result is that the voltage across  $C_2$  is twice the peak line voltage, and the voltage across  $C_3$  is equal to the peak line voltage. Assume, as an example, that the line voltage is 120 V rms. This will result in a voltage across  $C_2$  of 340 V, and a voltage across  $C_3$  of 170 V. The total voltage across the two capacitors (in series) is 510 V—which is three times the peak input line voltage. This voltage appears across the output terminals of the supply. The output voltage of the tripler is high, but the circuit cannot sustain a heavy load current.

If a negative output voltage is required, the diodes in any of the circuits can all be reversed. If this is done, the connections for the electrolytic capacitors must also be reversed because these capacitors are *polarized*—that is, they can only be connected into a circuit with the correct polarity.

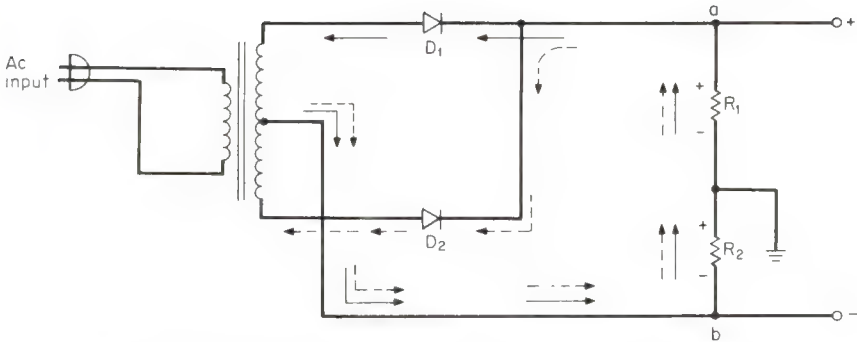
**LOAD VERSUS LOAD RESISTANCE** The terms load and load resistance have been used interchangeably in discussions up to now, but these terms actually mean different things. The load of a power supply always refers to the amount of *current* that the power supply must deliver. Therefore, when it is written that a power supply is under a *heavy load*, it means that the power supply must deliver a heavy current to the load. The load resistance, on the other hand, is the value of load resistance across the power-supply output terminals. The higher the load resistance value, the lower the amount of current that must be delivered by the power supply. Therefore, a high load resistance means a low power-supply load.

**COMBINATION POSITIVE AND NEGATIVE SUPPLIES** In some circuits it is desirable to obtain both a positive and a negative output voltage from the same power supply. For example, an amplifier may use both pnp and npn transistors, and a vacuum-tube amplifier requires a positive plate voltage and a negative grid voltage. Transformer-type power supplies can be readily adapted to provide both polarities from the same circuit. Such a circuit is shown in Fig. 16.11. Instead of using a common tie point at the center tap of the transformer secondary winding, this point is connected to one side of a series resistance branch comprised of  $R_1$  and  $R_2$ . The junction of the two resistors is tied to the common point of the circuit.

Since this is a full-wave rectifier circuit, the current will flow through the resistors on both half-cycles of input signal. The arrows show the paths of current flow. The polarities of the resulting voltage drops across the resistors are marked on the diagram. Since the current is flowing from point  $b$  toward ground, point  $b$  must be a negative voltage with respect to ground. At the same time current is flowing from the ground point toward point  $a$ , making point  $a$  positive with respect to ground. Both voltage polarities are obtained from the same power-supply circuit. The output voltages are usually filtered by two different filter circuits—one for each polarity of voltage.

If the circuit to which the power-supply voltage is being delivered requires a highly positive voltage for operation of the plates of a tube and a relatively low negative voltage for operating the grid, a separate transformer may be used to obtain the negative voltage. For transformerless power supplies, the positive voltage may be obtained with a half-wave rectifier or other transformerless supply, and the negative dc voltage





**Fig. 16.11** This circuit will supply a positive and a negative output voltage.

may be obtained by a separate transformer supply. The advantage of using two separate supplies over that of the circuit shown in Fig. 16.11 is that the negative supply can be more efficiently designed. Normally the grid circuits of vacuum-tube amplifiers do not draw current, and therefore the load on such a supply is small. A small, inexpensive transformer may be used for the negative grid circuit supply. The identical power-supply principles are equally applicable to power supplies for solid-state circuits.

### 16.3 REGULATED POWER SUPPLIES

An ideal regulated dc power supply is one that maintains a constant dc output, regardless of changes in external load requirements or changes in input voltage. The degree to which the performance of an actual power supply approaches this ideal is defined by its specifications, which will be discussed in depth later in this chapter. Figure 16.12 shows a general-purpose regulated power supply used in laboratories.



**Fig. 16.12** A general-purpose lab power supply. (Courtesy Lambda Electronics Corp.)

In general, dc power supplies may be divided into two functional categories: constant-current and constant-voltage types. Constant-current supplies maintain a constant current through a load resistor, regardless of the value of the load resistance, within a certain voltage range. Constant-voltage supplies maintain a fixed voltage across the load resistor as long as the current through the load resistor stays within a specified range. Equations (16.1 and 16.2) give the maximum and minimum values of load resistance for a constant current power supply.



$$R_{\max} = \frac{E_{\max}}{I} \quad (16.1)$$

$$R_{\min} = \frac{E_{\min}}{I} \quad (16.2)$$

where  $R_{\max}$  = maximum value of load resistance,  $\Omega$   
 $R_{\min}$  = minimum value of load resistance,  $\Omega$   
 $E_{\max}$  = maximum value of output voltage, V  
 $E_{\min}$  = minimum value of output voltage, V  
 $I$  = output current, A

**example 16.1** A constant-current power supply has an output current of one ampere and a voltage range (called the *compliance*) of 0 to 50 V. What are the maximum and minimum values of load resistance?

**solution** Using Eq. (16.1), we may find the maximum load resistance:

$$R_{\max} = \frac{E_{\max}}{I} = \frac{50 \text{ V}}{1 \text{ A}} = 50 \Omega \quad (16.1)$$

Using Eq. (16.2), we may find the minimum load resistance:

$$R_{\min} = \frac{E_{\min}}{I} = \frac{0 \text{ V}}{1 \text{ A}} = 0 \Omega \quad (16.2)$$

In this case, the minimum load resistance is zero ohms, which is a short circuit. Equations (16.3) and (16.4) give the maximum and minimum values of load resistance for a constant-voltage power supply.

$$R_{\max} = \frac{E}{I_{\min}} \quad (16.3)$$

$$R_{\min} = \frac{E}{I_{\max}} \quad (16.4)$$

where  $R_{\max}$ ,  $R_{\min}$  = same values as stated above

$E$  = output voltage, V

$I_{\min}$  = minimum value of output current, A

$I_{\max}$  = maximum value of output current, A

In most supplies,  $I_{\min} = 0$ ; the maximum load resistance for a constant-voltage supply is then infinity, or an open circuit.

**example 16.2** A constant-voltage power supply has an output voltage of 50 V and an output current range from 50 mA to one ampere. What are the maximum and minimum values of load resistance?

**solution** Using Eq. (16.3) to find the maximum load resistance,

$$R_{\max} = \frac{E}{I_{\min}} = \frac{50 \text{ V}}{50 \text{ mA}} = 1000 \Omega$$

Using Eq. (16.4) to find the minimum load resistance,

$$R_{\min} = \frac{E}{I_{\max}} = \frac{50 \text{ V}}{1 \text{ A}} = 50 \Omega$$

If this supply had been capable of going to a no-load condition, zero amperes output current, the maximum load resistance would be infinite or an open circuit.

Many power supplies are designed to operate in either a constant-voltage mode or in a constant-current mode. They are usually equipped with *automatic crossover*, a feature that allows the supply to go directly from one mode of operation to the other as the load resistance changes.

Many supplies are designed only for constant-voltage operation over a narrow range. These are frequently referred to as *slot-range* supplies. They generally are equipped with a *foldback* current-limiter circuit, which functions to decrease the output voltage and current simultaneously when the load resistance is too low; i.e., the output current rating of the supply would be exceeded if the output voltage remained at its normal value.

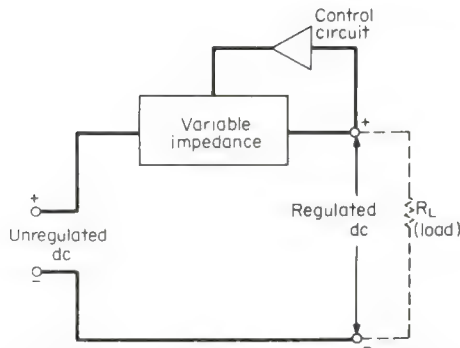


Fig. 16.13 Block diagram of a series-regulator power supply.

**BASIC TYPES** Regulated dc power supplies may also be classified by the type of input supply they utilize: ac or dc. The ac type usually employs a transformer to isolate the input from the output and to either raise or lower the input voltage to the level required. The ac output of the transformer is then rectified and filtered. The filter output is an unregulated dc voltage. From this point on, the regulator which follows may be considered a "dc-to-dc regulator." Thus, once voltage changing, rectification, and filtering are accomplished, the ac-to-dc and the dc-to-dc regulators may be treated alike.

There are three basic methods of regulating an unregulated dc source. Many power systems utilize a combination of two or in some cases all three of these.

**Series type** The most common system employed is the series regulator, which is shown in block form in Fig. 16.13. The series regulator is essentially a variable impedance placed between the unregulated input and the regulated output. By varying this impedance, the output voltage (or current) is maintained constant. The variable impedance is an active device, usually a power transistor in modern equipment, but vacuum tubes are utilized in older supplies.

**Shunt type** Another frequently encountered regulation system is the shunt regulator, shown in block form in Fig. 16.14. Here a fixed-source impedance is placed between the unregulated source and the regulated output. A variable impedance, called a "shunt" or "parallel" impedance, is placed across the output. The shunt impedance changes to maintain the output voltage (or current) constant. As in the case of the series regulator, the variable impedance is an active device, either a transistor or a vacuum tube.

**Switching type** In either of these two above approaches, all energy not delivered to the load must be dissipated in the regulating elements. More efficient systems involve switching circuits which utilize some form of *duty-cycle* control. The details

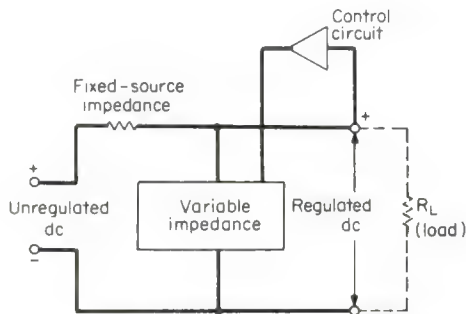


Fig. 16.14 Block diagram of a shunt regulator.

of this *nondissipative* type of regulator will be discussed in more detail later in this chapter. Some systems use only a nondissipative circuit, while others employ this circuitry as a coarse preregulator followed by a fine regulator of either the shunt or series type. This produces the finer regulation of the dissipative control, while having the efficiency advantage of a "nondissipative" circuit.

#### 16.4 POWER-SUPPLY REGULATOR CIRCUITS

Figure 16.15 shows a simple power-supply regulator consisting of a series resistor  $R_s$  and a zener diode  $D_1$ . A gaseous regulator tube would serve the same purpose as the zener diode in this circuit, but it is not available with low-voltage ratings. It usually takes more than 70 V to "fire" a gaseous regulator tube, while the zener has the advantage that it is available in a wide selection of voltage values. Another advantage of the zener diode over the gaseous regulator is that it is available in a much wider range of current-handling abilities.

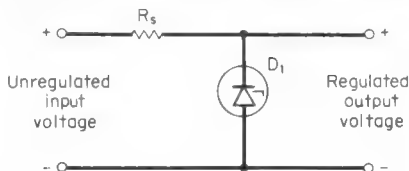


Fig. 16.15 A simple power-supply regulator.

The value of resistance for the series resistor in the simple shunt regulator of Fig. 16.15 can be calculated from the equation

$$R_s = \frac{E_i - E_o}{I_m + I_z} \quad (16.5)$$

where  $E_i$  = unregulated input voltage, V

$E_o$  = regulated output voltage, V

$I_m$  = maximum load current, A

$I_z$  = current through the zener diode, A

As a general rule, the zener diode current  $I_z$  should be made equal to about 10 percent of the value of the maximum load current. Mathematically,

$$I_z = 0.1I_m \quad (16.6)$$

where  $I_m$  and  $I_z$  have the same meaning as for Eq. (16.5).

The power rating of  $R_s$  can be determined from the equation

$$P_s = 2(I_m + I_z)^2 R_s \quad (16.7)$$

Once the zener current rating is obtained from Eq. (16.7), then its power rating can be determined from the equation

$$P_z = E_o I_z \quad (16.8)$$

Manufacturer's specifications can be consulted in order to choose the zener that meets the particular requirements of a power supply. A sample problem will show how the regulated zener supply can be designed.

**example 16.3** Design a 14-V regulated supply that will provide a maximum load current of 100 mA from an unregulated power supply that delivers 25 V.

**solution** A 14-V zener diode will be needed.

The zener current rating:

$$\begin{aligned} I_z &= 0.1I_m \\ &= 0.1 \times 100 \text{ mA} = 10 \text{ mA} \end{aligned} \quad (16.6)$$

Resistance value of  $R_s$  from Eq. (16.5):

$$R_s = \frac{E_i - E_o}{I_m + I_z} = \frac{25 - 14}{0.1 + 0.01} = 100 \, \Omega$$

Power rating of  $R_s$  from Eq. (16.9):

$$P_s = 2(I_m + I_z)^2 R_s = 2(0.1 + 0.01)^2 \times 100 = 2.42 \text{ W}$$

$$P_z = E_o I_z = 14 \times 0.01 = 0.14 \text{ W}$$

The series regulator will have a resistance of  $100 \Omega$  and a power rating of over  $2.4 \text{ W}$ . The  $14\text{-V}$  zener will have a current rating of  $10 \text{ mA}$  and a power rating of over  $0.28 \text{ W}$ .

Figure 16.16 shows a closed-loop power-supply regulator circuit. At the unregulated input side of the regulator, there is a series circuit comprised of  $R_1$ ,  $Q_1$ , and zener diode  $D_1$ . At the regulated output terminals there is a series resistor circuit comprised of  $R_2$ ,  $R_3$ , and  $R_4$ .

Resistor  $R_2$  is a variable resistor that sets the base voltage of  $Q_1$ . The series resistor string is sometimes called the *sense circuit* because it delivers any change in the output voltage to the base of transistor  $Q_1$ .

In order to understand the operation of the power-supply regulator shown in Fig. 16.16, it will be assumed that the regulated output voltage begins to rise for some reason (such as a decrease in load current). The emitter voltage of  $Q_1$  is a fixed value set by the zener diode. Therefore, when the output voltage rises, it causes the base of  $Q_1$  to become more positive, causing  $Q_1$  to conduct harder through resistor  $R_1$ . The base voltage for control amplifier transistor  $Q_2$  is obtained from the drop across  $R_1$ . Increasing the current through  $R_1$  causes the base of  $Q_2$  to become less positive. This decreases the current through  $Q_2$ , which controls the base current of  $Q_3$ . Decreasing the base current of  $Q_3$ , the power regulator transistor, has the same effect as increasing the resistance between the emitter and collector of the transistor.

To summarize, a rise in the voltage at the output terminals is accompanied by an increase in resistance at  $Q_2$ , causing an increased drop across  $Q_3$  and bringing the voltage at the output terminals back to its regulated value.

A decrease in the output voltage causes the base at  $Q_1$  to become less positive. Therefore, it conducts less through resistor  $R_1$ . That causes the base of  $Q_2$  to go in a positive direction, increasing the current through  $Q_2$  and also through  $Q_3$ . This is equivalent to decreasing the resistance of  $Q_3$  and decreasing the voltage drop across it.

To summarize, a decrease in voltage at the output terminals is accompanied by a decrease in drop across  $Q_3$ , and the output voltage remains constant. In this closed-loop feedback regulator circuit,  $Q_3$  serves as a variable resistor which is automatically adjusted to compensate for changes in load voltage. The voltage across  $Q_3$  increases or decreases, as required, in order to maintain the voltage across the output terminals at a constant.

The closed-loop regulator of Fig. 16.16 will also regulate against changes in input voltage. Suppose, for example, that the input voltage—that is, the unregulated voltage

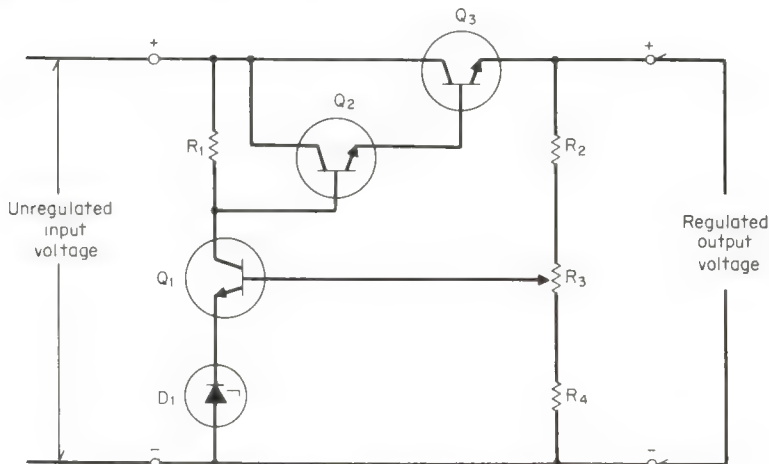


Fig. 16.16 A closed-loop regulator.

—increases. This makes the collector voltage of  $Q_1$  greater and causes it to conduct harder. The base of transistor  $Q_2$  becomes less positive, decreasing the current through that transistor and hence decreasing the current through  $Q_3$ . This is the same as increasing the emitter-to-collector resistance of  $Q_3$  and increasing the voltage drop across it. Thus, an increase in the input voltage on the unregulated side is accompanied by an increase in the voltage drop across  $Q_3$ , and the output voltage at the regulated side does not change.

A similar chain of events occurs when the output voltage decreases. The result is that when the output voltage decreases, the feedback loop sets the base of  $Q_3$  so that it conducts harder. This is the same as lowering the resistance of  $Q_3$  and decreasing the voltage drop across it. Thus, when the input voltage decreases, the decrease also occurs across  $Q_3$ , and the voltage remains unchanged.

# Chapter 17

## Batteries

### 17.1 INTRODUCTION

The use of chemical reactions to produce electricity dates from March 1800 when the Italian physicist, Alessandro Volta (1745–1827), described two batteries to the Royal Society of London. The first of these was called a “Crown of Cups” since it consisted of a group of cups arranged in a circular pattern. Each cup was filled with salt solution (the electrolyte) and contained two strips (electrodes) of dissimilar metals (silver and zinc), with each of the two strips externally connected to its adjoining mate. These, of course, were series-connected and were the forerunner of our various modern cylindrical cell systems. In the other design that Volta presented, the dissimilar metals were stacked in the form of a “pile,” very much like coins in a dispenser, with a piece of paper soaked in salt separating alternate layers of metal. This was the forerunner of our modern-day wafer or flat-plate cells.

For the most part, batteries representing variations of the Crown of Cups concept were used until about 1860 when Gaston Plante described a battery using lead and lead oxide electrodes and an electrolyte of sulfuric acid to the French Academy of Science. This, of course, ultimately developed into our familiar automobile battery, and has by now evolved into a myriad of sizes and engineering types to serve a very large number of applications. Plante’s contribution was particularly noteworthy in that it achieved high-power levels.

Working at about the same time, another Frenchman, Georges LeClanche, conceived the idea of having all the battery’s electrolyte absorbed by its electrodes so as to essentially function as though it were dry. In 1868 LeClanche received a patent for such a “dry” cell and undertook its manufacture. As one of the electrodes in his cell, LeClanche used manganese dioxide and a carbon-rod current collector. A cup of zinc not only formed the cell container, but also served as the other electrode. LeClanche’s cell, therefore, was the direct lineal forebear of our familiar flashlight dry cells, and was a most significant development since it provided a truly mobile source of electric energy, thereby triggering the growth of the battery industry and a host of battery-using devices.

Today we have a broad spectrum of devices that yield electric energy from a controlled chemical reaction. In size, these range from tiny mercury “button” cells such as are used to power hearing aids, to the large lead-acid cells used to provide all power requirements, including propulsion of submerged submarines. A small button cell (i.e., the M5 size) will be less than  $\frac{1}{8}$  in. in diameter by a little more than  $\frac{1}{8}$  in high and will weigh about 0.02 oz. A perfectly conventional submarine cell will be about 18 in wide, 18 in long,  $5\frac{1}{2}$  ft high, and will weigh over one ton. The little M5 has a capacity of 36 mAh whereas its giant submarine cell cousin has a capacity measured in kilowatt-hours.\*

\* The abbreviations “h” for “hour” and “s” for “second” follow the SI system recommended by the IEEE.



## 17-2 Batteries

Not only is there an almost continuum of sizes between these two extremes, but there are now, as regularly available commercial products, a number of different cell systems, each having a combination of properties against which specific needs can be matched. In addition to Plante's cell and LeClanche's cell, we can consider nickel-cadmium batteries for extended cycle life, nickel-iron cells for ruggedness, silver-zinc cells for very high power, mercury cells for very high capacities, alkaline cells for outstanding versatility, magnesium dry cells for stability in hot, humid climates, and so on. In addition to sizes and types, additional variety of operating properties is obtained from different basic physical forms. Cells are available in cylindrical designs, rectilinear forms (using flat electrodes or tubular electrodes). Other forms include wafer cells (both round or rectangular) for easy series stacking for high voltages, and now even in pile form for still easier stacking to still higher voltages. In spite of this variety, the underlying principles whereby chemical reactions are used to provide electric energy—or the process is reversed so that electric energy can be used to bring about chemical reactions—can be readily understood.

**GENERAL CONSIDERATIONS** Of the various kinds of chemical reactions that can take place, one type generally called a redox reaction (i.e., reduction-oxidation) is of particular interest to us. A redox reaction not only involves a chemical reaction between two substances, but also a transfer of electrons from one of the substances (the reducing agent) to other substances (the oxidizing agent). As might be expected, if a redox reaction can be set up in such a manner as to cause the transfer of electrons to take place through a wire, we would in effect have a means of generating an electric current. Just such conditions are set up in a cell or battery which thereby acts as a source of electricity. (A battery is a group of cells connected in series and parallel to provide the desired voltage and power. By common usage, single cells are also commonly referred to as batteries.)

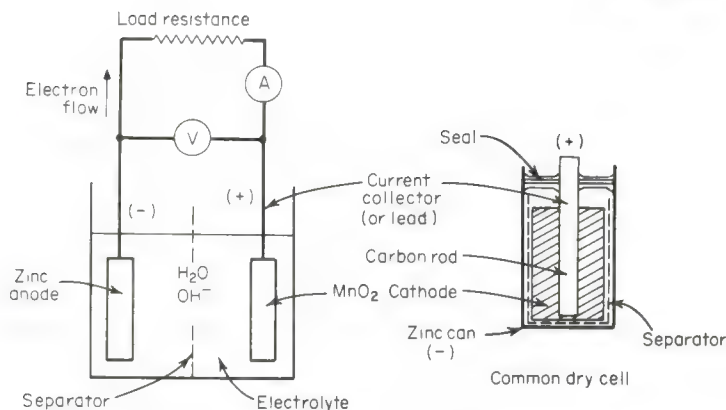
There are three essential components in a battery: an anode, a cathode, and an electrolyte. The anode and cathode together are referred to as the electrodes and are made up of the chemicals which will ultimately react to cause the electron transfer. In addition, the electrodes must provide a means for physically supporting the reacting chemicals, and a suitable conductive path for facilitating the flow of the electrons. This latter structure is referred to as the grid of the electrode. The anode is the electrode which will release electrons (and become oxidized) during the discharge of the battery. Similarly the cathode is the electrode which will absorb electrons (and be reduced) during discharge of the battery. The electrolyte is a liquid, usually an aqueous solution made conductive by the dissolution of an acid, a base, or a salt. The electrolyte completes the conductive path between the electrodes of opposite polarity. We can summarize and illustrate this with a conventional primary dry cell.

**PRIMARY BATTERY** In the common dry cell (also referred to as a carbon-zinc or LeClanche cell), manganese dioxide ( $\text{MnO}_2$ ) is used as a cathode and zinc (Zn) as an anode. The electrolyte is a solution of ammonium chloride and/or zinc chloride. When the anode and cathode are connected through a wire so as to permit electrons to flow from anode to cathode, the reaction between  $\text{MnO}_2$  and Zn can take place, and an electric current is set up in the external circuit and the battery is discharging, as illustrated in Fig. 17.1.

In operation, zinc is converted to zinc oxide ( $\text{ZnO}$ ), releasing two electrons which then flow through the external circuit to become available to the  $\text{MnO}_2$ . These electrons are then absorbed by the  $\text{MnO}_2$ , thereby changing the latter to a lower oxide of manganese (such as  $\text{Mn}_2\text{O}_3$ ). It should be noted that hydroxyl ions ( $\text{OH}$ ) and water also enter into the reactions, but their net participation is canceled out. (Hydroxyl ions are particles in solution made up of hydrogen and oxygen atoms and carry a single negative charge.) As can be seen, the battery can continue to set up an electron flow until one of the other chemical reactants is exhausted, in which case the battery is said to be discharged.

In summary, a battery stores chemicals (not electricity) and can bring about a reaction between these chemicals in such a manner as to cause an electron flow through an external circuit.

The cell we have just described is classified as a primary battery since the chemical



**Fig. 17.1** Fundamental operation of a dry (primary) battery.

Anode reaction:



Cathode reaction:



Overall reaction (obtained by adding anode and cathode reactions):

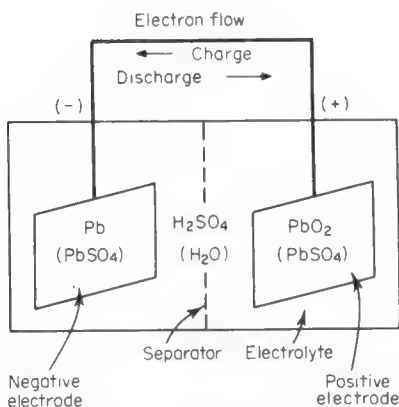
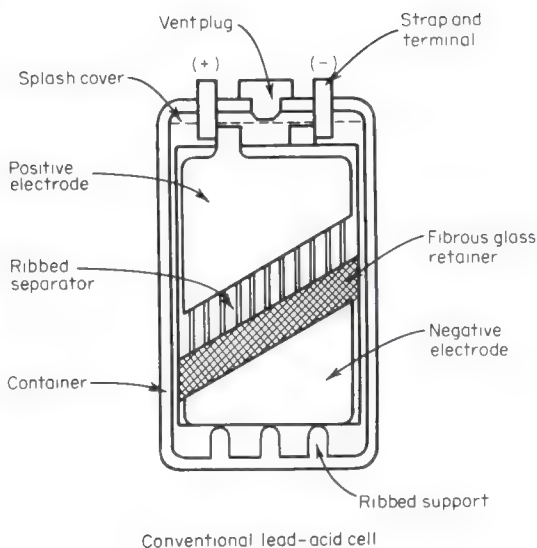


reactions that gave rise to the electric current are not readily reversible. By convention, the electric current flow is opposite to that of the electron flow, and the cathode is therefore designated as the positive electrode, and the anode as the negative.

**SECONDARY BATTERY** Once the active materials are used up, a primary battery is not capable of yielding any more electric energy and is then discarded. Other battery systems use materials that can be restored to their original chemical state by reversing the current flow, i.e., by providing electric energy to the cell from some external source. This process is known as charging, and a battery that is capable of undergoing a number of discharge-charge cycles is called a secondary battery. An example of a secondary battery is the lead-acid system, illustrated in Fig. 17.2. By definition, the anode on discharge becomes the cathode on charge. Similarly, the cathode on discharge becomes the anode on charge. However, the terminal designations of positive and negative do not change since the current flow reverses when the roles of the electrodes of secondary batteries are most often referred to as positives and negatives, and while this designation can also be used for primaries, the electrodes of the latter are more usually referred to as anodes and cathodes.

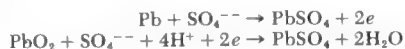
In their most common forms both primary and secondary batteries are entirely operative when they are made. Because of this, slow deteriorations can occur owing to a small amount of spontaneous reaction or "local action" that goes on all the time in activated batteries.

**RESERVE BATTERIES** Where extra long shelf life (i.e., ability to yield its electric energy after long storage) is desired, the electrolyte is withheld from the battery, and steps are taken to exclude moisture since water, in some form, is usually required for the local action reactions. Such batteries are extremely stable in storage, and are activated by the addition of the appropriate electrolyte. Since the power of such batteries is thereby kept in reserve, they are frequently referred to as "reserve batteries." A special type of reserve battery is sometimes also referred to as a "dry charged" battery. In this case the electrodes of the battery are brought to a state of full charge at the time they are made, and then deactivated by the removal of electrolyte. The restoration of electrolyte at the time and place the battery needed results in a virtually

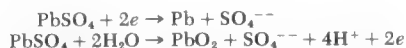


**Fig. 17.2** Fundamental operation of a lead-acid (secondary) battery.

Discharge:



Charge:



factory-fresh battery. A great many automobile batteries are now shipped in the dry charged condition with the station attendant adding the (sulfuric acid) electrolyte at the time of installation.

There is a special class of reserve batteries that use nonaqueous or molten salt electrolytes. Such electrolytes contain no water and utilize organic liquids, or salts that become conductive in the molten condition. This latter kind of electrolyte must be liquefied before it becomes functional, and the battery can therefore be activated by

heat. This is usually accomplished by a heat squib, which melts the electrolyte to activate the battery. Such batteries are known as thermal-type reserve batteries, heat-activated reserve batteries, or simply thermal batteries. A typical thermal battery will be in the form of a small button cell with a calcium or magnesium anode, a silver chromate cathode, and an electrolyte consisting of a mixture of potassium, sodium, and lithium chlorides. Such batteries have an indefinite storage capability, but discharge times of only a few minutes at best. However, they have found special uses such as power for proximity fuses. Further descriptions of reserve batteries will be considered later in this chapter along with other special types.

## 17.2 CHARACTERISTICS OF IMPORTANT COMMERCIAL PRIMARY BATTERIES

A large number of different primary-cell types are now being produced commercially, but only relatively few have reached the point where they are in volume production and widely distributed. Actually, each cell system has physical and electrical properties different from all the others, so that the user now has an extensive choice from which to select the type most adaptable to his particular application. Table 17.1 shows the comparative properties of the different primary-cell systems that are most widely distributed and of commercial importance.

**TABLE 17.1 Characteristics of Commercially Important Primary Cells**

Type	Open-circuit voltage, V	Average operating voltage, V	Wh/lb	Wh/in <sup>3</sup>
LeClanche	1.50-1.65	1.25	5-40	0.5-3
Alkaline	1.52	1.20	20-50	2.0-3.5
Mercury	1.35-1.40	1.25	25-55	5-8
Magnesium dry	1.90-1.95	1.50	15-50	1-4

Of these, the most important is the LeClanche system, also known as carbon-zinc cells or simply "dry cells." This latter designation has come into widespread use although the cells are certainly not dry. As a matter of fact, they contain electrolyte just as any other battery does, but this electrolyte is absorbed into the cathode active material so that the cell appears to be dry.

**BASIC TYPES** The dry-cell industry has standardized on three basic engineering types: round cells, flat cells, and wafer cells. The industry has further standardized dimensions for 14 round cell sizes, 4 flat cell sizes, and 5 wafer sizes. Virtually all batteries are made up from these standard sizes, detailed in Table 17.2. Many manufacturers, however, issue their own catalog information showing the exact dimensions of their finished cells and batteries. Dry cells are called upon to provide a wide variety of electrical needs in terms of capacity, voltage, and current. For example, a flashlight may require 0.25 A at 3 V, while a small transistor radio may require only 5 mA at 9 V. To satisfy an almost endless variety of capacity-voltage-current requirements, the industry has developed these standard sizes and types, which are then combined in series-parallel arrangements to satisfy almost any desired power requirement.

**POLARIZATION OF DRY CELLS** During the discharge of a dry cell, the electrolyte actively enters into the electrode reactions. In the vicinity of the actual reaction sites, the electrolyte at one electrode becomes depleted of its reactants, while the electrolyte at the other electrode becomes concentrated with the products of the reaction. This condition, known as polarization, slows down (or may even block) any further reaction. This has the effect of lowering the voltage at which the battery is actually discharging, or restricting the magnitude of the current that can be taken from the battery. When the battery is allowed to stand on open circuit, the natural diffusion of the components of the electrolyte permits the reactants and products to redistribute themselves more uniformly, thereby reducing (or even entirely overcoming) the effects of polarization.

After this occurs, the battery can continue its discharge. Thus, dry cells can yield

TABLE 17.2 Standard Commercial LeClanche (Dry) Cell Sizes

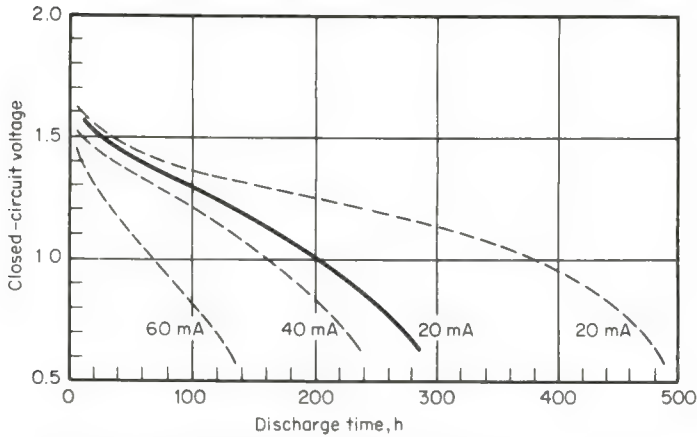
Cell designation		Nominal dimensions <sup>a</sup> , in		
American National Standards Institute	International Electrochemical Commission	Diameter	Height of can	
ROUND CELLS				
N	R1	$\frac{27}{64}$	$1\frac{1}{64}$	
N	R1	$\frac{7}{16}$	$1\frac{3}{32}$	
R	R4	$\frac{17}{32}$	$1\frac{3}{8}$	
AAA	R03	$\frac{25}{64}$	$1\frac{1}{16}$	
AA	R6	$\frac{17}{32}$	$1\frac{7}{8}$	
A	R8	$\frac{5}{8}$	$1\frac{7}{8}$	
B	R12	$\frac{3}{4}$	$2\frac{1}{8}$	
C	R14	$\frac{15}{16}$	$1\frac{13}{16}$	
CD	R18	1	$3\frac{9}{16}$	
D	R20	$1\frac{1}{4}$	$2\frac{1}{4}$	
E	R22	$1\frac{1}{4}$	$2\frac{7}{8}$	
F	R25	$1\frac{1}{4}$	$3\frac{7}{16}$	
G	R26	$1\frac{1}{4}$	$3\frac{13}{16}$	
6	R40	$2\frac{1}{2}$	6	
		Length	Width	Thickness
FLAT CELLS				
F40	F40	$1\frac{1}{4}$	$\frac{27}{32}$	$\frac{7}{32}$
F70	F70	2	$1\frac{1}{4}$	$\frac{7}{32}$
F90	F90	$1\frac{5}{8}$	$1\frac{9}{16}$	$1\frac{13}{16}$
F100	F100	$2\frac{13}{32}$	$1\frac{13}{16}$	$1\frac{27}{32}$
WAFER CELLS				
F15	F15	$\frac{9}{16}$	$\frac{9}{16}$	$\frac{1}{8}$
F20	F20	$\frac{15}{16}$	$\frac{17}{32}$	$\frac{1}{8}$
F30	F30	$1\frac{1}{4}$	$\frac{27}{32}$	$\frac{1}{8}$
F25	F25	$\frac{15}{16}$	$\frac{15}{16}$	$\frac{7}{32}$
F40	F40	$1\frac{1}{4}$	$\frac{27}{32}$	$\frac{9}{32}$

<sup>a</sup> Bare cells, i.e., less manufacturer's finishing and labeling.

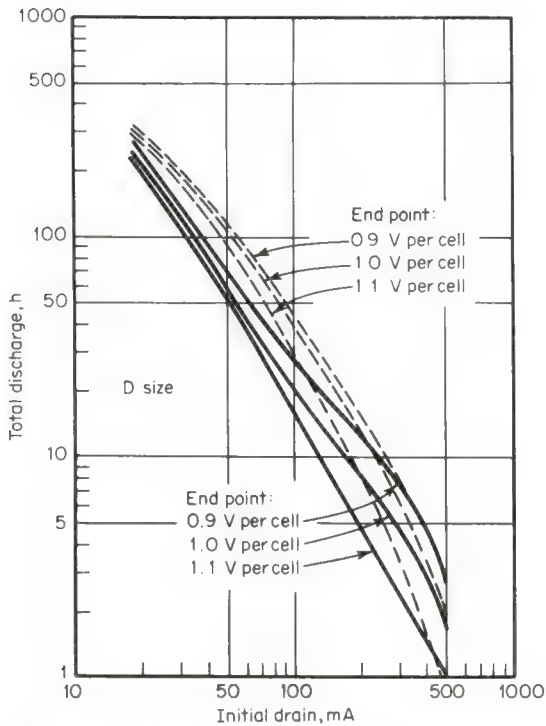
more energy on an intermittent discharge routine than on a continuous one. As can be expected from the foregoing, a normal discharge of a dry cell shows a continuously decreasing voltage until the desired cutoff voltage (usually 0.9 V) is reached. Since diffusion is working against polarization, cells discharging at low rates (where diffusion can keep pace with polarization) show very slowly decreasing voltages, dropping off rapidly, however, as the battery is exhausted. Cells discharging at high rates (where polarization effects dominate) show very rapidly decreasing voltages during discharge, but good recovery after a "rest" period. This is illustrated in Fig. 17.3.

**PERFORMANCE DATA** Because of the wide variety of circumstances under which dry cells are discharged, performance data are difficult to summarize. Nevertheless, some fairly good methods of correlating the anticipated performance with the use conditions have been developed. One fairly widely used procedure is to correlate the capacity, expressed as operating time, with the initial drain rate at the start of the discharge, i.e., after initial stabilization of battery voltage. Such discharges are based on the load having a constant resistance. A graphical presentation usually gives the service life in hours as the ordinate, against an abscissa of initial drain rate in milliamperes. For convenience, a log-log plot is used. As indicated earlier, minimum serviceable voltage also has an effect, and such plots usually show a family of curves for several end of discharge voltages. Figures 17.4 and 17.5 present representative rating curves for a round cell and a wafer cell, operating continuously and intermittently.



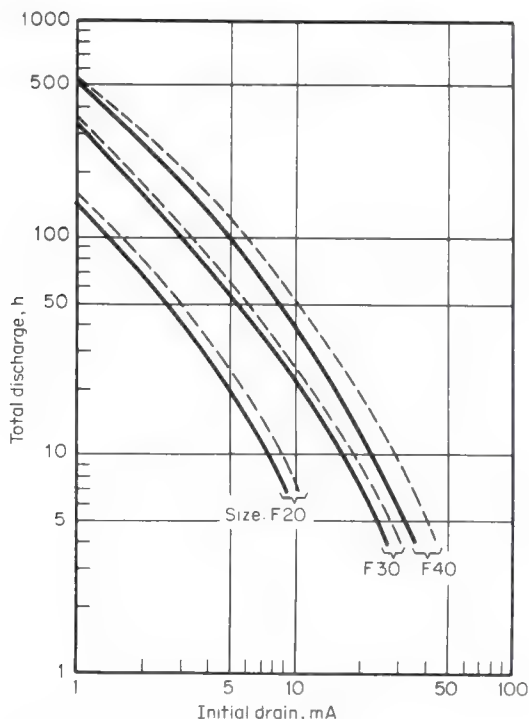


**Fig. 17.3** Discharge characteristics of LeClanche dry cells; D-size cells showing continuous discharge (solid line) and high and low rates of intermittent discharges for 4 h/day (dashed lines).



**Fig. 17.4** Rating curves for typical round LeClanche cell. Solid line shows continuous discharge; dashed line intermittent discharge (4 h/day).





**Fig. 17.5** Rating curves for typical wafer dry cells. Solid lines show continuous discharge; dashed lines show intermittent discharge (4 h/day).

**example 17.1** How long can a D cell be expected to continuously operate a resistive load of  $25\ \Omega$  to an end point of 0.9 V?

To approximate the initial drain, note that the initial stable cell voltage will be about 1.5 V. (See Fig. 17.3.)

$$I = \frac{E}{R} = \frac{1.5}{25} = 0.06\ \text{A} \\ = 60\ \text{mA}$$

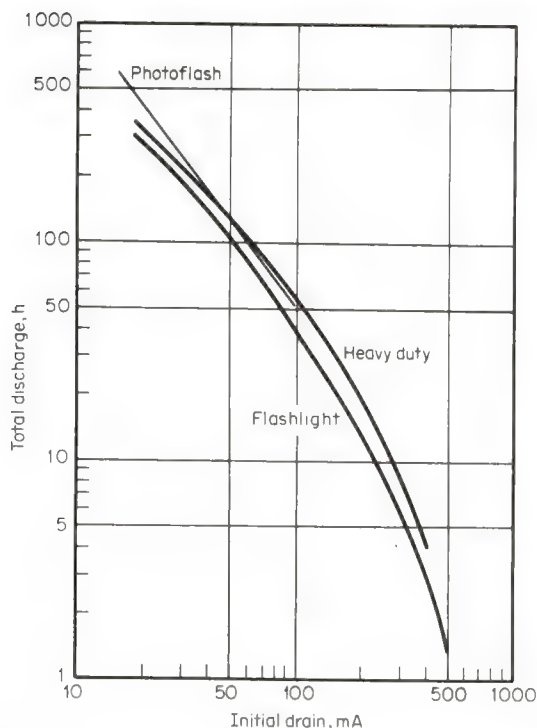
On Fig. 17.4, locate an initial drain of 60 mA on the abscissa (horizontal axis). Follow this point vertically upward until it intersects the topmost solid curve (which represents a 0.9-V end point) and read 53 h as the ordinate (vertical value) of that point.

**example 17.2** How long can the cell of Example 17.1 be expected to operate if used intermittently for four hours per day to the same end point?

For this problem, follow the 60-mA initial drain line until it intersects the topmost dashed line, and read its ordinate as 90 h.

The manufacturers of dry batteries have considerable control over the performance of their batteries through control of the mix formulations. Those batteries intended for essentially low drain use, but requiring high capacity (as for example in powering transistor radios) use mixes designed to give these properties; others needing only low capacity, but at high drain rates (as for example in photoflash applications) use mix formulations that give this combination. This is illustrated in Fig. 17.6 showing how discharge characteristics can be controlled even in a single physical size.

**example 17.3** How much longer will a flashlight having an initial drain of 200 mA operate on a heavy-duty D than on a general-service flashlight battery if both are operated to a 1.0-V/cell end point?



**Fig. 17.6** Rating curves illustrating effect of mix composition; D-size cells showing intermittent discharge (4 h/day), low voltage percent.

Using Fig. 17.6, find the initial drain of 200 mA on the abscissa, and follow it vertically upward until it intersects the flashlight battery line and read 15 h as the ordinate. Then continue vertically until the 200-mA line intersects the heavy-duty battery line and read 20 h as the ordinate. The heavy-duty battery will therefore yield  $20 - 15 = 5$  h of additional service.

Another means of specifying capacity is in terms of the amount of electricity that can be drained under the conditions of discharge. This is usually stated in ampere-hours, milliamper-hours, or ampere-minutes. Capacity expressed in this way, however, is sometimes difficult to arrive at since with a constant load the current will vary with time as a result of the battery voltage varying with time. Under such conditions, the ampere-hour capacity can be taken as the product of the average current and the discharge time.

**AMERICAN NATIONAL STANDARDS INSTITUTE REQUIREMENTS** For many purposes, the American National Standards Institute has developed test procedures and minimum specified requirements. They are described in detail in National Bureau of Standards Handbook 71, but some of the more important test procedures and specifications are abstracted in Table 17.3.

**example 17.4** A radio A battery uses 3F-size dry cells in series. How shall it be tested for compliance with ANSI Standards? Does this battery meet ANSI Standards if it runs 41 days on this test?

Using Table 17.3, the test shown for radio A service is set up to operate the battery daily for one four-hour period each day. The load will be  $3 \times 25 = 75 \Omega$  to an end point of  $3 \times 1.0 = 3.0$  V. If the battery runs 46 days, it will have yielded  $41 \times 4 = 164$  h. Since the ANSI requirement is 140 h, this battery is well over the minimum requirement.

TABLE 17.3 ANSI Test Procedures and Minimum Performance Specifications

Test name	Daily discharge schedule	Resistance per $1\frac{1}{2}$ -V unit, $\Omega$	End point per $1\frac{1}{2}$ -V unit, $\Omega$	Cell size	ANSI initial requirement
General-purpose flashlight to represent 0.5-A lamp	1.5-min period	2.25	0.65	D general purpose	400 min
General-purpose flashlight to represent 0.3-A lamp	1.5-min period	4	0.75	C AA	325 min 80 min
General-purpose flashlight to represent 0.22-A lamp	1.5-min period	5	0.75	AAA	50 min
Heavy industrial flashlight to represent 0.3-A lamp	32 4-min periods	4	0.9	D industrial	800 min
Light industrial flashlight to represent 0.3-A lamp	8 4-min periods	4	0.9	D general purpose D industrial	600 min 950 min
Railroad lantern to represent 0.15-A lamp	8 $\frac{1}{2}$ -h periods	8	0.9	F	45 h
Photoflash test	60 1-s periods (1 discharge/min for 1 h/day)	0.15	0.5 D size, 0.25 C & AA sizes	D photoflash C photoflash AA photoflash	800 s 700 s 150 s
Radio A	1 4-h period	25	1.0	F G	140 h 170 h
Radio B	1 4-h period	$166\frac{2}{3}$	1.0	F40 A F90	30 h 130 h 225 h
Heavy intermittent	2 1-h periods	$\frac{2}{3}$	0.85	F (4 cells in parallel) #6 general purpose #6 industrial	70 h

**SHELF LIFE** Among the special characteristics of dry batteries, a very important one is shelf life. As previously discussed, batteries produce their electric energy as a result of reactions that take place simultaneously at two otherwise independent electrodes. Under certain circumstances it is possible for each electrode reaction to take place independently of the other, thus not providing useful electric energy to an external circuit. Such reactions are referred to as "local action" phenomena, and have the effect of using up the materials without providing useful electric energy. In a battery, local action predominately affects the anode or negative electrode. Usually one of the by-products of local action is the generation of hydrogen gas which can result in a buildup of internal pressure, and some provision must therefore be made to allow hydrogen to escape. This is usually done by making the central carbon electrode porous so that hydrogen can slowly diffuse safely from the cell. It should be noted that this represents a design difficulty, since water which must be kept in the cell in order for the cell to operate properly, can also escape by this route (particularly at elevated temperatures).

Through developmental efforts over the years, the shelf life of quality dry batteries in casual storage (i.e., at normal temperatures and humidities) has reached the 18 to 21 months level, at which time the battery can still give up to two-thirds of its original service life.

In considering the storage of LeClanche cells, low temperatures are very beneficial. For example, a dry cell stored at 32°F will retain virtually all its original capacity for as long as two years. It should be noted, however, that the battery should be allowed to warm up to room temperature before it can yield its full capacity. This should be done slowly to prevent condensed moisture from physically disrupting the battery assembly.

The LeClanche system is also capable of designs that produce high voltages in small spaces. For such batteries, the flat-type cell or wafer-type cell is most frequently used. These batteries find application where high capacities are not required, but high voltages are. Since the flat cell can be readily stacked in rectangular spaces, the efficiency of space utilization for such batteries is high. In addition, the series connection of a large number of cells to produce the required voltages can be efficiently handled. Flat cells or wafer-cell batteries have therefore found extensive use in instrumentation, as radio B batteries and even radio A batteries (particularly where the power requirements are not excessive).

The special electrical properties and variety of available designs and shapes have resulted in the very wide use of LeClanche batteries in flashlights, radio receivers and transmitters, photographic equipment, portable meters, and other (generally low-power) devices.

**ALKALINE BATTERIES** At first appearance, alkaline-MnO<sub>2</sub> primary cells, some times called alkaline dry cells, or simply alkaline cells, seem to be very closely related to LeClanche cells. In point of fact, there are a great many similarities such as the electrochemically active materials which are the same (MnO<sub>2</sub> cathode, zinc anode) as illustrated in Fig. 17.7. In addition, the similarity is continued in available physical

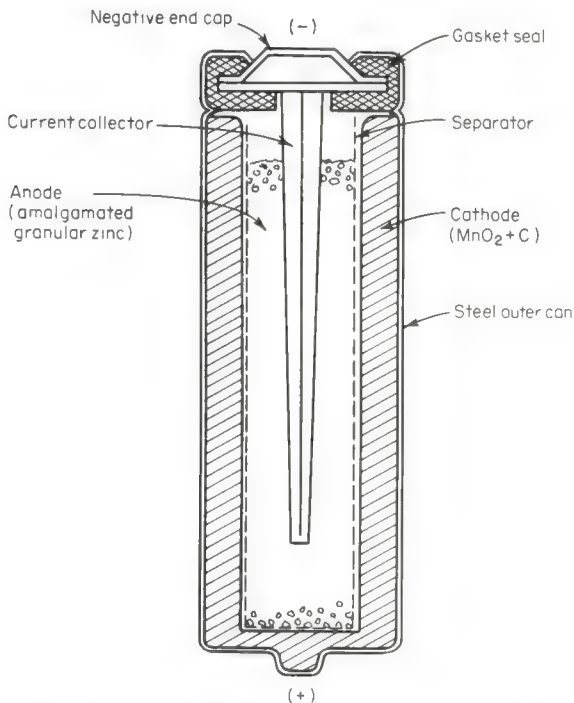


Fig. 17.7 Schematic diagram of a typical alkaline primary cell.

sizes and shapes. However, in the alkaline cell, a solution of a strong alkali, potassium hydroxide, is used as the electrolyte in place of the  $\text{ZnCl}_2$  and  $\text{NH}_4\text{Cl}$  of the conventional LeClanche system. In addition, the zinc exists in granular form so that it presents a very extensive surface to the electrolyte. These factors greatly alter the electrical and other characteristics of the battery. Polarization effects in alkaline cells, for example, are much less pronounced than in conventional dry cells (under equivalent discharge conditions), and much of the energy can be withdrawn at higher power levels. Along related lines, it is also found that their discharge characteristics are more easily defined and catalogued since the voltage variation during discharge is likely to be somewhat less than in a LeClanche cell. Alkaline dry batteries are, therefore, rated according to their capacity, usually cited at a particular drain rate. For convenience the initial drain rate is used, and the capacity is usually expressed in milliampere-hours or ampere-hours at this rate. Table 17.4 summarizes the sizes and electrical characteristics of the commercially important alkaline cell sizes.

**TABLE 17.4 Sizes and Characteristics of Alkaline Primary Cells**

Size designation	Capacity		Diameter, <sup>o</sup> in	Height, <sup>o</sup> in
	mAh	At initial drain of mA		
N	580	18	0.47	1.12
AAA	750	18	0.41	1.75
AA	1,800	60	0.55	1.94
C	5,000	100	1.31	1.93
D	10,000	300	1.31	2.37

<sup>o</sup> Finished cells; i.e., including manufacturer's labels, etc.

**example 17.5** How long can an alkaline AA cell be expected to operate at a drain of (a) 60 mA, (b) 40 mA, (c) 100 mA?

**solution** From Table 17.4, the capacity of an AA alkaline cell is 1800 mAh at a drain of 60 mA. The operating time therefore will be  $1800/60 = 30$  h. At 40 mA the capacity will be slightly higher, but for estimating purposes would still be taken as 1800 mAh:  $1800/40 = 45$  h. At 100 mA the capacity will be slightly lower, but for estimating purposes can still be taken as 1800 mAh:  $1800/100 = 18$  h. (At higher current drains, the capacity should be determined for the specific use and circumstances, the capacity given in the table being used only for approximations.)

Like LeClanche cells, alkaline cells have generally good shelf-life characteristics. As a matter of fact, alkaline cells are actually sealed since local action reactions in healthy cells are not serious. As a consequence, hydrogen production is very much slower than in LeClanche cells, and in fact by proper care during manufacture can be made negligible. Since alkaline cells are sealed, there is also no loss of moisture during storage, and alkaline cells therefore accept prolonged storage very well. As a matter of fact, a shelf life of two to three years (to 80 percent of original capacity) is common. Also, because they are sealed, alkaline cells can be stored at higher temperatures than LeClanche cells. Storage up to 130°F for extended periods of time is entirely feasible. However, since elevated temperatures accelerate local action reactions, it is not recommended that they be exposed to temperatures above 130°F for any length of time. In the other direction, as the temperature of the cells is reduced, there is only a slow falloff in capacity for alkaline cells, making them well-suited for operation in moderately low-temperature ambients. Quantitatively the effects of temperature are summarized in Table 17.5.

Although the alkaline dry cell is generally not troubled with the evolution of hydrogen gas, under adverse conditions such as overdischarging or excessive temperatures some gassing can occur, and most alkaline cells are therefore provided with a safety blowoff which permits hydrogen to be vented at a certain pressure so that the hydrogen can be harmlessly relieved.

Because of their basic and special properties and ability to provide relatively high power levels, alkaline dry batteries are now extensively used to power portable radios (particularly where high power levels are required), in photographic equipment, and



**TABLE 17.5 Effect of Ambient Temperature on Capacities of Alkaline Primary Cells**

Discharge temperature, °F	Approximate Service, % °		
	Light drain	Medium drain	Heavy drain
113	100	100	100
70	100	100	25
30	70	40	25
-4	25	20	5

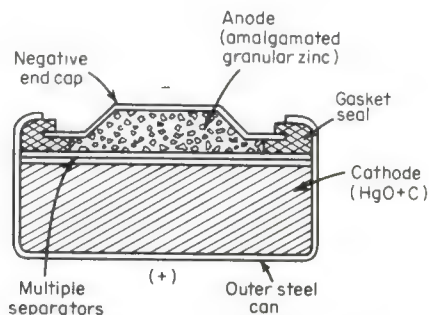
*Note.* When discharge times (at normal ambient temperatures) exceed approximately 30 h, drains are considered light, when less than approximately 15 h, drains are considered heavy.  
 ° Cells cooled to ambient temperature before discharge; service at 70°F to 0.8-V end point taken as 100%.

for driving fractional horsepower electric motors, such as are used in electric shavers, hobbycraft, and the like.

**MERCURY DRY BATTERIES** Another cell system using an alkaline electrolyte that has achieved a significant commercial position is the mercury dry cell. The outstanding characteristics of mercury cells is that they represent the highest available capacities for a given volume of any of the primary batteries. As can be seen in Fig. 17.8, one type of mercury battery is very similar in construction to the alkaline battery. Physically the major difference is that mercuric oxide ( $\text{HgO}$ ) is substituted for the manganese dioxide, but this brings about some profound changes in cell characteristics. In addition to having very high capacities, mercury cells also have an outstanding shelf life since mercury or mercury compounds have a strong effect in reducing local action. As a matter of fact, the shelf-life problem as far as mercury cells are concerned is the ability to thoroughly seal the cell rather than any chemical action that might take place.

Because of this, the practical shelf life seems to be limited to about two or three years on larger sizes, and perhaps to one or two years on the smaller sizes. Mercury cells are marketed in essentially the same sizes as alkaline and LeClanche dry cells, but in addition are available in extremely small sizes called button cells, schematically illustrated in Fig. 17.8. Table 17.6 gives the general sizes and weights of the more important commercial sizes.

The temperature characteristics of mercury cells are very similar to those of alkaline cells, except that at low temperatures capacities fall off a little more rapidly, so that the lowest temperature suitable for mercury-cell operation would be about 30°F, as shown in Table 17.7.



**Fig. 17.8** Schematic diagram of typical mercury button cell. Note that cylindrical mercury cells closely resemble alkaline primary cells, shown in Fig. 17.7, differing principally in the composition of the cathode mix.



## 17-14 Batteries

**TABLE 17.6 Sizes of Commercially Important Mercury Cells**

Size designation, ANSI	Rated capacity		Nominal dimensions, in. <sup>o</sup>	
	mAh	At drain of mA	Diameter	Height
M5	36	2	0.31	0.14
M8	100	3	0.46	0.13
M15	160	5	0.46	0.21
M25	350	3	0.46	0.57
M40	1000	35	0.63	0.65
M55(AA)	2400	50	0.55	1.97
M100(D)	14000	250	1.28	2.39

<sup>o</sup> Bare cells, i.e., less manufacturer's finishing and labeling.

**TABLE 17.7 Temperature Characteristics of Mercury Cells**

Discharge temperature, °F	Approximate service, % <sup>o</sup>	
	Light drain	Heavy drain
113	100	100
70	100	100
40	95	7
30	6	2

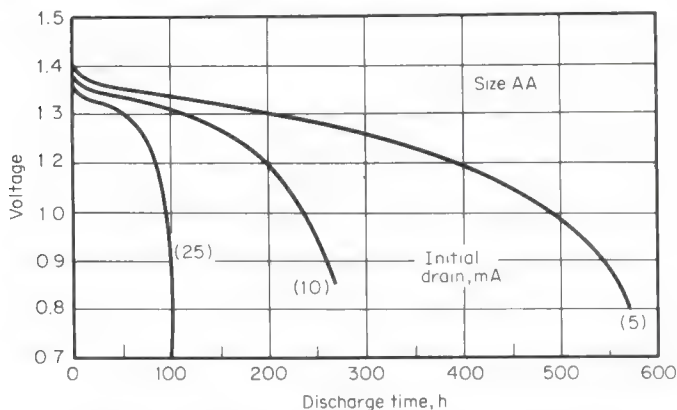
*Note.* When discharge times (at normal ambient temperatures) exceed approximately 25 h, drains are considered light; when less than approximately 25 h, drains are considered heavy.

<sup>o</sup> Cells cooled to ambient temperature before discharge, service at 70°F to 0.9-V end point taken as 100%.

**example 17.6** A mercury battery can operate a particular radio receiver for 160 h at normal ambients. How long can it be expected to operate the receiver in an ambient of 35°F?

Table 17.7 indicates that the battery would operate the radio (i.e., a light drain) for 0.95(160 = 152 h at 40°F, and only 0.06(160) = 9.6 h at 30°F. To estimate performance at 35°F, it is necessary to extrapolate between these values:  $0.5(152 - 9.6) + 9.6 = 80.8$  h.

In addition to their extremely good shelf life and high capacity, mercury cells have another outstanding characteristic: namely, uniformity of discharge voltage. As can be seen from Fig. 17.9, mercury cells have an almost uniform discharge voltage, except for a short time at the very beginning and of course at the very end of the discharge.



**Fig. 17.9** Typical discharge characteristics of mercury cells.

This permits designers of instrumentation and electronic equipment to use circuits that need not accommodate voltage fluctuations. Not only is the voltage of mercury cells stable, but it is also highly reproduceable. Mercury cells, therefore, find wide application in instrumentation as stable voltage working cells (to be periodically matched against standard cells for very high precision and long service). As a consequence of their high capacity, availability in very small sizes, good shelf life, and stable discharge voltage, mercury cells have found a wide range of applications, as, for example, in portable voice recorders, radio transceivers, and the like.

**MAGNESIUM DRY BATTERIES** No discussion of primary batteries would be complete without significant mention of the new magnesium dry cell. Fundamentally, the magnesium dry cell is an exact counterpart of the LeClanche system in which the outer can is made of magnesium alloy instead of zinc.

This rather "small" change is accompanied by a number of more subtle changes that result in an entirely new system with a set of characteristics all its own. The outstanding properties of the magnesium dry battery are its ability to withstand high-temperature, high-humidity storage, and extremely long shelf life under casual storage. In addition, the magnesium dry battery operates at voltage levels considerably higher than conventional LeClanche or mercury dry cells. This is due to the fact that magnesium is a far more electropositive (i.e., active) metal than zinc. However, because of its reactivity, one would ordinarily expect a magnesium cell to be unstable, but what actually happens is that magnesium forms a very thin but very impervious oxide coating, thereby protecting the underlying metal from local action. Although special electrolytes must be used (generally solutions of magnesium bromide or magnesium perchlorate) and the cell must be closed by very special techniques in order to retain moisture and still permit hydrogen to escape (since hydrogen is liberated as a by-product of the discharge), it is really the protective film which forms on the magnesium that has made the magnesium dry cell possible. At their present stage of development, these batteries can be successfully stored at 160°F for 30 days without significant loss of capacity (i.e., less than 10 percent).

As yet, the total production from all suppliers is being funneled into military applications, but it is anticipated that units will soon be available for normal commercial applications. Cell sizes have not as yet been standardized, and only cursory performance data can be presented at this time. What general information is available is presented in Table 17.8 and Figure 17.10.

**TABLE 17.8 Sizes of Typical Magnesium Cells**

Cell designation, ANSI	Nominal dimensions, in°	
	Diameter	Height
N	0.42	1.09
R	0.53	1.38
A	0.63	1.88
CD	1.00	3.19
D	1.25	2.25

° Bare cells, i.e., less manufacturer's finishing and labeling.

As can be seen from Table 17.1, magnesium dry batteries are fully in a position to compete with mercury and alkaline batteries in terms of power capability per unit weight and capacity per unit volume. Ultimately, magnesium dry cells should also be more favorably priced than their mercury or alkaline counterparts. At least for the immediate future, however, on the basis of costs, magnesium dry cells are not expected to make serious inroads into conventional dry battery areas, except for those applications where extremely long shelf life (even up to five years) is required, or where the batteries are to be used in hot, humid areas where the logistics of getting fresh LeClanche cells to the user are poor. Magnesium batteries, since they can withstand extremely adverse storage conditions for extended periods of time, can be delivered

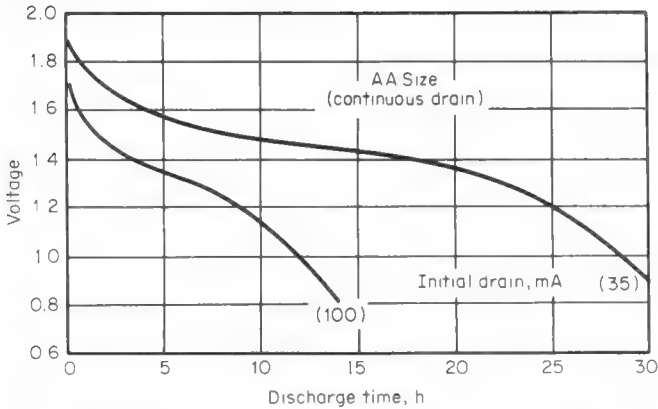


Fig. 17.10 Typical discharge characteristics of magnesium dry cells.

to the ultimate user in tropical and semitropical countries at full or almost full capacity. The outstanding shelf life of the system (as compared to that of conventional dry cells) is illustrated by Fig. 17.11. Here both the carbon-zinc and magnesium types show very little capacity loss after three months' storage at room temperature, but for storage at elevated temperatures the superiority of the magnesium system quickly asserts itself. At really high storage temperatures, the carbon-zinc type fails rather quickly, while the magnesium type retains most of its original capability.

There are other primary batteries that have achieved a degree of commercial value, such as silver-zinc, air-depolarized, and others, but these are not yet in sufficiently general use to warrant discussion at this time. However, it should be mentioned that the silver-zinc primary does have an outstanding high rate capability and a capacity as much as 80 percent greater than its secondary counterpart, described later in this chapter. Physically it is very similar to the dry charged form of the secondary version (shown in Fig. 17.15) except that it usually does not have an ion-restrictive diaphragm. The higher capacity is due to the fact that the silver electrode can be oxidized to a higher level before cell assembly than can conveniently be done after cell assembly and activation (as would be the case for secondary operation).

Before going on to rechargeable systems, Fig. 17.12 is included to illustrate the actual physical forms of some of the primary systems we have considered.

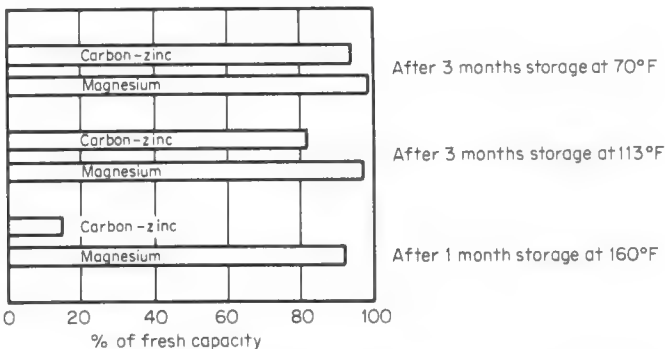


Fig. 17.11 Comparison of high-temperature shelf-life capabilities of carbon-zinc and magnesium dry batteries.



Fig. 17.12 Sample display of primary batteries. (Courtesy Burgess Battery Company)

### 17.3 CHARACTERISTICS OF IMPORTANT COMMERCIAL SECONDARY TYPES

Although there are a large number of types among the secondary, or rechargeable batteries, only three have achieved any really significant commercial importance. These are the lead-acid, nickel-cadmium, and silver-zinc systems.

**LEAD-ACID BATTERY** Of these, the lead-acid system is the most important from a commercial point of view and is widely distributed. It is available in a very large variety of sizes and shapes to satisfy applications requiring a broad range of capacities and power levels. A summary of commercially important types is given in Table 17.9. It should be noted that while the battery sizes are arranged in order of increasing capacity, some are for 6-V (three-cell), 12-V (six-cell), or 2-V (single-cell) units so that the table covers a much broader range than at first appearance.

An interesting point should be noted concerning the 50- and 100-Ah units which are approximately the same size and weight in spite of the twofold capacity difference. This is explained by the fact that they both have the same energy content, 600 Wh (obtained by multiplying the capacity in ampere-hours by the battery voltage, as indicated above). It is important to realize that the capacity and power capability of a rechargeable battery depend in very large measure on its prior history. A typical situation, particularly in lead-acid batteries, is that the capacity builds up during early cycle life,

TABLE 17.9 Representative Sizes and Weights of Lead-Acid Batteries

Nominal size Ah capacity @ 20-h rate	Number of cells	Nominal dimensions, in°			Approximate weight
		Length	Width	Height	
4	3	2 <sup>3</sup> / <sub>4</sub>	1 <sup>7</sup> / <sub>8</sub>	4	1 <sup>3</sup> / <sub>4</sub>
8	3	5 <sup>1</sup> / <sub>2</sub>	2	4 <sup>3</sup> / <sub>4</sub>	4
30	6	8	5 <sup>3</sup> / <sub>8</sub>	9 <sup>1</sup> / <sub>8</sub>	27
50	6	11 <sup>1</sup> / <sub>2</sub>	7 <sup>1</sup> / <sub>8</sub>	9 <sup>1</sup> / <sub>8</sub>	50
100	3	10 <sup>3</sup> / <sub>8</sub>	7 <sup>1</sup> / <sub>8</sub>	9 <sup>1</sup> / <sub>8</sub>	46
200	6	21 <sup>5</sup> / <sub>8</sub>	11 <sup>1</sup> / <sub>8</sub>	10 <sup>1</sup> / <sub>4</sub>	147
280	1	5 <sup>1</sup> / <sub>8</sub>	6 <sup>1</sup> / <sub>4</sub>	15 <sup>1</sup> / <sub>4</sub>	60
420	1	5 <sup>1</sup> / <sub>8</sub>	6 <sup>1</sup> / <sub>4</sub>	21 <sup>3</sup> / <sub>8</sub>	84

\* Overall dimensions, i.e., including terminals and vent plugs.

then maintains a substantially uniform value for most of the battery's useful life, and finally falls off (more and more rapidly) toward the end of life. Unless some accidental life-terminating event occurs, such as a short circuit or loss of electrolyte through a cracked case, etc., the end of life of the battery will be reached when its capacity falls below that which is useful for the application. For a great many applications, this will be at about 60 percent of the original capacity.

As for the rating of new batteries, different manufacturers hold to somewhat different views. Some, for example, might rate a battery on what it will do on the very first discharge cycle in the user's hands, whereas others may take advantage of the fact that the capacity normally builds up during early cycle life, and will rate a battery after it has been "conditioned" in use. Where the capacity of the battery is critical, the user should determine which of these systems is the basis for rating the battery.

The conditions of discharge also have a very profound effect on the capacity and power capability of rechargeable batteries. As a general rule, a battery will display its greatest capacity when discharged at very low rates. For lead-acid batteries this rate is generally chosen to be the 20-h rate. A great deal of confusion has arisen as a result of the industry method of expressing the discharge rate of a battery. The phrase "the 20-h rate" means that the battery is being discharged at such a current as will exhaust the capacity of the battery in 20 h. The confusion results from the fact that it is necessary to know what the capacity of the battery is at the particular current drain in order to determine the current at which it should be discharged.

To simplify this situation, the discharge current is now generally related to a nominal capacity of the battery. For example, if the nominal capacity of a battery is 10 Ah, a discharge rate of 0.5 A would be called the "20-h rate." Most leading producers now cite capacity at the 20-h rate and at normal (78°F) ambient temperatures. For lead-acid batteries, there is usually very little difference in capacity as the rate of discharge is increased from the 20-h rate to perhaps the 10-h rate. But thereafter, as the discharge rate is further increased, there is an increasingly rapid falloff of capacity. Except for very special designs, the half-hour rate is generally considered to be the maximum sustained rate at which lead-acid batteries can be used with any degree of efficiency.

The lead-acid system has a number of characteristics that are particularly noteworthy. Although these batteries do not withstand high-temperature storage very well (anything above 130°F being somewhat deleterious), they do withstand low-temperature storage extremely well and are capable of yielding considerable power and operation at low temperatures. As a typical example, the capacity of a lead-acid battery is likely to drop off only about 30 percent as the temperature is decreased from room temperature to 0°F. Additional information is given in Table 17.10. In determining the capacity of lead-acid batteries from Table 17.10, it should be noted that the discharge rate has considerable significance. In using the table, the column calling for the discharge rate at which the battery is to be operated as measured at room temperature must first be found, and this column is then followed downward to the appropriate temperature. The figure cited is the capacity (expressed as a percent of the 20-h rate capacity at room temperature) at the same current drain.

**TABLE 17.10 Effect of Temperature on Capacity of Lead-Acid Batteries**

Temperature, °F	Running time, hours <sup>o</sup>					
	1	5	10	50	100	500
	Drain current, amperes					
100	118	21	10	1.4	0.6	0.06
80	114	20	9.0	1.3	0.5	0.05
60	106	19	8.3	1.1	0.5	0.05
40	97	17	7.5	1.0	0.4	0.04
20	90	15	6.7	0.9	0.3	0.03
10	82	13	5.6	0.8	0.2	0.02

<sup>o</sup> Capacity at 80°F and 20-h rate = 100 Ah.



**example 17.7** A battery has a capacity of 100 Ah at the 20-h rate at room temperature. What is its capacity at the 10-h rate at room temperature and at 20°F?

**solution** For a 100-Ah battery the current drain at the 20-h rate is 5A. From Table 17.10, its capacity at room temperature at the 10-h rate will be 91 percent of 100 Ah, or 91 Ah. This is the nominal capacity of the battery at room temperature and at the 10-h rate. For this battery, therefore, the discharge current at room temperature and the 10-h rate is 9.1 A. Following the 10-h rate column downward to 20°F, the capacity at a drain of 9.1 A will be 66 percent of what it was at room temperature, or 60 Ah. This table may seem a little difficult to use, but it is a very serviceable one in that it contains a great deal of information since interpolations for other rates and temperatures give entirely acceptable results.

In a rechargeable battery, charge retention, i.e., the ability to maintain capacity after charge, corresponds to shelf life in a primary battery. In a conventional lead-acid battery (sometimes referred to as an antimonial alloy lead battery), the charge retention varies considerably from the early stages of life to the latter stages, falling off slowly but continuously as the battery is cycled. When new, the battery is very likely to be capable of losses of less than 10 percent per month in (casual) storage, but this figure could easily approach 100 percent for an old unit nearing the end of its useful life. A newer type of lead-acid battery is now available, however (i.e., the calcium alloy lead battery), which has a remarkable degree of charge retention that does not fall off to any significant extent as the battery is used. In a calcium alloy type, the charge retention is likely to be better than 85 percent after as long as one year in casual storage, and this will be fairly uniform throughout its life. Generally speaking, this latter type is used primarily for standby power since it has only a limited cycle life capability.

Because of the variety of properties that can be developed by design and/or electrical treatment, lead-acid batteries have achieved an amazing variety of applications. Apart from the popular ones of providing starting, lighting, and ignition power for automobiles, operating electric fork-lift trucks and similar commercial applications, they have now found significant applications in standby power (for telephones, industrial purposes, hospitals, emergency lighting, etc.), as well as in a variety of special applications such as portable TV, electric power tools, and ordnance guidance systems, to name but a few of the more important ones.

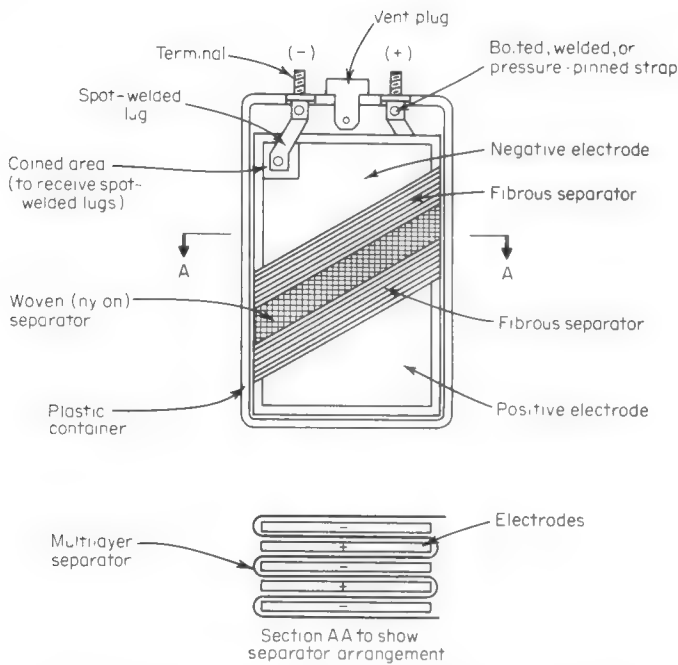
**NICKEL-CADMIUM BATTERIES** A second type of rechargeable battery of great and continuously growing commercial importance is the nickel-cadmium system. Within the basic system, there are several different engineering types and designs, each design reflecting the development of specific properties for a specific field of application. The pocket-type nickel-cadmium battery, for example, is an extremely long-life unit capable of thousands of very deep discharges. Because it can also moderate rates of discharge, it is used principally where motive power is desired, such as in electric wagons and the like. Of more recent development, the sintered-plate-type nickel-cadmium battery is available in both the sealed and vented forms. The sintered-plate system is by far the fastest growing type in the nickel-cadmium field and will be discussed in more detail.

Basically, the sintered-plate type utilizes a very porous nickel to support the active materials and serve as a current path. This matrix is made by sintering finely divided particles of carbonyl nickel powder (a special grade of "macelike" particles). The resulting plaques can be as much as 80 percent porous and are impregnated with the active materials, nickel hydroxide in the positive electrode, and metallic cadmium in the negative. As is common in all nickel-cadmium types, the sintered-plate nickel-cadmium battery utilizes a potassium hydroxide electrolyte. A schematic diagram of a sintered-plate nickel-cadmium battery is shown in Fig. 17.13, and common commercially available sizes are given in Table 17.11.

Vented cells are usually encased in plastic containers, while sealed cells are usually in thin-walled steel containers. Both are extremely rugged, and the battery can take a good deal of physical abuse, as well as electrical abuse, without permanently deleterious effects. Nickel-cadmium batteries are capable of very high rates of discharge, and as a matter of fact many applications will discharge such batteries at the one, and even the one-half hour rate at good efficiencies. The high rate capability of these batteries also makes them useful for "pulsed" discharges (i.e., very high rate, very short duration discharges).

In addition to their fine physical characteristics, nickel-cadmium batteries perform





**Fig. 17.13** Schematic diagram of sintered-plate nickel-cadmium cell.

**TABLE 17.11** Sizes of Common Nickel-Cadmium Cells

ANSI size designation	Nominal capacity		Size, in		
	mAh	At drain of mA	Diameter	Height	
SEALED, ROUND					
AA	50	5	0.61	0.24	
	150	15	0.98	0.28	
	450	45	1.70	0.30	
	450	45	0.56	1.95	
C	1200	120	0.88	1.67	
D	4000	400	1.30	2.43	
Capacity			Size, in		
	Nominal Ah	At drain of amperes	Length	Width	Height
SEALED, PRISMATIC					
	1.5	0.15	1.36	0.75	3.36
	4.5	0.45	1.93	1.67	3.13
	7.5	0.75	1.93	1.67	4.25
	15	1.50	3.56	1.34	4.90
VENTED, PRISMATIC					
	6.5	0.65	0.96	2.18	4.06
	24	2.4	1.08	3.16	8.28
	40	4.0	1.39	3.16	9.40
	120	12.0	2.17	6.81	8.73
	230	23.0	3.17	8.14	9.40

well at both high and low temperatures, but capacity does fall off if the temperature is raised or lowered from ambient room temperatures. The temperature characteristics of typical nickel-cadmium cells are shown in Table 17.12. As can be seen from Table 17.12, the falloff in capacity as the temperature is lowered is about the same as for lead-acid types for the temperature range covered. However, as the temperature is lowered beyond the range covered in Tables 17.10 and 17.12, the performance of the nickel-cadmium battery falls off more slowly than its lead-acid counterpart. Here only qualitative data can be cited (since we are in an area where specific designs play a most important role), but a nickel-cadmium battery can be quite useful down to  $-40^{\circ}\text{F}$  (i.e., yielding about 5 to 10 percent of its normal capacity), whereas a lead-acid battery would probably have a practical limit of  $-20^{\circ}\text{F}$ . There is another very important factor to consider as regards high rate capability at low temperatures, in that the current capability per unit weight or per unit volume, for short-duration discharges, is very favorable for nickel-cadmium sintered-plate batteries. For these reasons this type of battery is favored for low-temperature engine-starting applications.

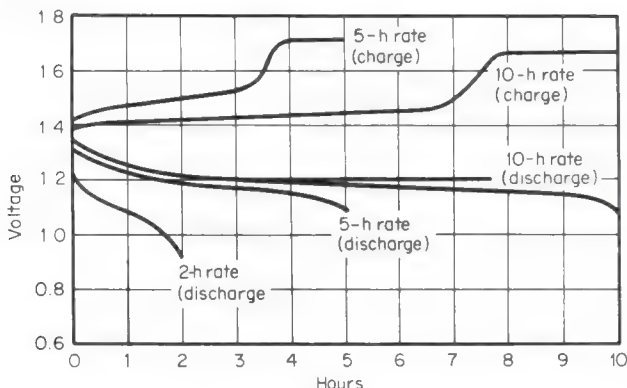
**TABLE 17.12 Temperature Characteristics of Nickel-Cadmium Cells**

Discharge temperature, $^{\circ}\text{F}$	% of room-temperature capacity at 10-h rate
115	93
78	100
40	93
32	90
0	60
-5	50

The charge retention characteristics of a nickel-cadmium battery are not too far different from those of a lead-acid battery, but in at least one respect the nickel-cadmium battery does have an outstanding characteristic: it can withstand extremely long periods of idle storage (up to 7 to 10 years being not at all uncommon) without permanent, deleterious effects. In fact, when it is desired to store a battery, it should be completely discharged all the way down to zero volts (by even short-circuiting the battery at low rates when its voltage has dropped to a point where this can be done safely). By contrast, a lead-acid battery must be electrically maintained during idle storage (by a monthly low-rate trickle, or some other routine-freshening charge). A nickel-cadmium battery needs no such electrical maintenance, and when placed into service after long idle storage will, after a few cycles, pretty much return to its proper capacity.

Perhaps one of the most outstanding characteristics of the nickel-cadmium battery is its ability to accept a recharge very quickly. Since the back emf of the battery rises relatively slowly during charge, as indicated on Fig. 17.14, most of the charge can be completed at high rates before the back emf begins to seriously limit the charge acceptance. However, since the battery is capable of accepting charges at very high rates, it is not overly important to control the charge to a great degree, particularly at the early stages. This is true for vented cells to a much greater degree than for sealed cells since the latter can also accept high overcharge rates. During overcharge of vented cells, however, water is electrolyzed into hydrogen and oxygen, using up water and electricity, but otherwise having no further deleterious effect on the battery.

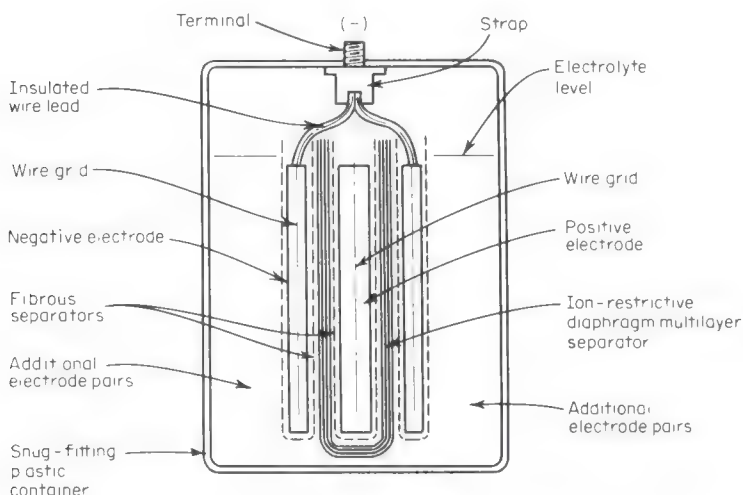
The nickel-cadmium system is thus a sophisticated combination of virtues and limitations. In spite of its high initial cost, when used under proper circumstances it can be quite economical since it provides the closest approach the battery industry has yet made toward the goal of long-life low (or no)-maintenance batteries. These batteries have therefore found extensive use in jet aircraft starting, radio and telephone communications, subcutaneous low-power medical applications, cold weather starting, low power starting (as in lawnmowers, chain saws, etc.), and, of course, in cordless appliances (such as toothbrushes, shavers, etc.).



**Fig. 17.14** Constant-current charge-discharge characteristics of nickel-cadmium batteries.

**SILVER-ZINC BATTERIES** In the spectrum of battery power sources, the silver-zinc system has achieved a firmly entrenched position principally because of its power capability. Whereas nickel-cadmium batteries do quite well at discharge rates as high as the one-hour rate, silver-zinc primaries and secondaries easily go to the ten-minute rate, and for pulse power even to the one-second rate with relatively good efficiencies. The general construction of silver-zinc secondaries is shown in Fig. 17.15. Essentially the primary version is made in about the same way, but the ion-restrictive diaphragm need not be used, nonwoven fabrics being used instead (to provide lower resistance and more rapid activation). Apart from stability in the dry condition (making it suitable for reserve battery applications), one of the principal virtues of the primary version is that it can take advantage of higher degrees of oxidation of the silver (since the positive electrode is "formed" before cell assembly under conditions favorable to silver oxidation). This manifests itself as a greatly increased discharge capacity.

Typical primary and secondary characteristics are shown in Fig. 17.16. This figure shows the voltage-time curves for a charge and discharge cycle of a silver-zinc secondary, and for the discharge of a primary. It is interesting to note that Fig. 17.16 makes



**Fig. 17.15** Schematic diagram of typical silver-zinc secondary cell (end view to show separator configuration).

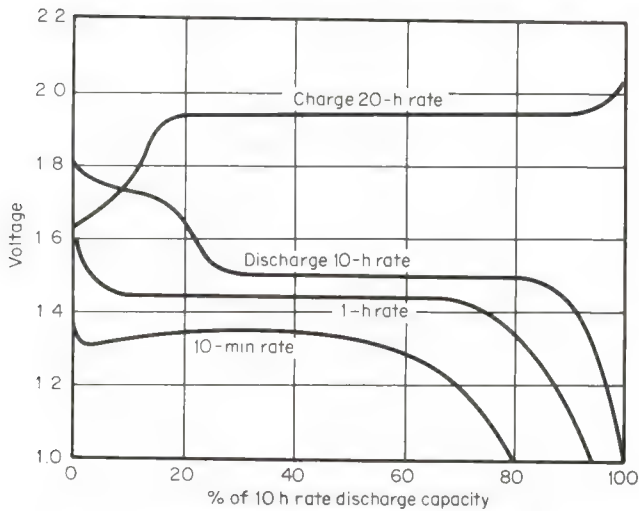


Fig. 17.16 Charge-discharge characteristics of typical silver-zinc secondary cell.

no reference to size (i.e., capacity) since the general shape of the curves fit all sizes with remarkable uniformity when each is operated at its appropriate current value. It is also interesting to note that while secondary batteries are usually furnished dry and uncharged, they can be furnished dry charged, in which case the first discharge will have the characteristics of a typical primary, reverting back to normal secondary characteristics for subsequent cycles. Of course, because of the presence of an ion-restrictive barrier, which usually is slow wetting, the activation time of a dry charged secondary will be as long as for any other silver-zinc secondary.

There are, of course, too many physical forms that have been developed for secondary batteries, but a representative sampling is shown in Fig. 17.17.



Fig. 17.17 Sample display of secondary batteries. (Courtesy Burgess Battery Company)

Referring now to the secondary version, the silver-zinc system shows a marked buildup in capacity during early (perhaps 20 percent of) life, followed by a long period (perhaps 70 percent of life) of reasonably stable capacity. Toward the end of life there is a rapid decline in capacity to a normal end point, usually about 60 percent of rated capacity. In operation the cycle life is highly variable, depending on use circumstances, but generally the range is from 20 for deep, rapidly repeated cycles, to perhaps 200 cycles of moderate (50 percent) depth, provided that sufficient time is allowed between discharge-charge-discharge to avoid temperature buildup.

High temperatures can be particularly harmful since the solubility of the negative electrode in the electrolyte is greatly increased, and deleterious degradation of the diaphragm also occurs. Overcharge also has seriously deleterious effects, but as in calcium-alloy batteries, a pronounced increase in back emf at the approach of the full-charge condition permits convenient control of the charge so that serious overcharging can be readily avoided.

Since silver oxide is one of the active materials, the cost of silver-zinc batteries is obviously high, limiting its applications to date to essentially military gear and special communication devices. Nevertheless it has an extremely good low-temperature capability, a modest charge retention (although this has been and is being improved), and a reasonable cycle life. For those applications requiring very high power levels, or extremely high bursts of energy (pulse loads) silver-zinc batteries indeed provide a convenient, reliable power source in small, lightweight packages.

#### 17.4 MAINTENANCE

Unlike primary batteries which require no special attention, the maintenance of secondary batteries is of extreme importance, particularly where maximum performance and life are to be obtained. Charging has a profound effect in this regard, and several good methods have been developed. The constant-current method simply returns the capacity that has been discharged at a fixed current for a fixed period of time, or until the back emf (i.e., the countervoltage) of the battery rises to some predetermined value. It is therefore sometimes called the time-controlled or voltage-controlled method of charging. This is perhaps the simplest method in terms of the required control equipment, but does not account for, or take advantage of the fact that the battery has the ability to accept a charge very rapidly at the early stages of recharge (particularly after a deep discharge) and that this situation is reversed as the battery approaches the fully charged condition.

As the battery approaches full charge, the efficiency with which the battery accepts the charge falls off very rapidly (reaching perhaps only 4 to 5 percent at about 95 percent of full charge). This results in an undesirable condition in constant-current charging, since it is too slow at the beginning of the recharge (when the battery can accept charge efficiently) and too fast toward the end of recharge (when the battery can accept charge only slowly). A wasteful recharge not only manifests itself as an undesirable cost, but also as excessive positive grid corrosion (which will result in short life) and excessive watering frequency.

To improve this situation, we can take advantage of the fact that the back emf of the battery rises as the state of charge increases. Many manufacturers now market charging devices which use a modified constant-potential procedure in which the initial stages of the recharge are carried out at a high, but constant current, the value of which is essentially dependent on the power capability of the charging source. As the back emf of the battery reaches a predetermined point, which is toward the latter part of the recharge, automatic switching circuits change the system to a fixed potential recharge. This potential is set so that it is a little higher than the end of charge back emf of the battery. As the back emf of the battery approaches the potential of the charging circuit, the charging current falls off as the ability of the battery to accept the charge falls off. This turns out to be efficient, not only in terms of time and cost, but also in the avoidance of overcharge and in reducing watering frequency.

**LEAD-ACID BATTERY MAINTENANCE** Apart from the selection of a suitable charging procedure, the electrolyte in lead-acid batteries must be maintained at the proper concentration and level. Negative electrodes exposed to the atmosphere as a result of



low electrolyte level are likely to oxidize very rapidly, a process that is only partially reversible by charging, and results in varying degrees of permanent damage to the electrode. When the concentration of sulfuric acid is allowed to fall, the capacity of the battery also suffers rather quickly since sulfate ions are a participant in the discharge reactions. In addition, the negative plates tend to sulfate (i.e., form an irreducible lead sulfate) in low-gravity acid.

On the other hand, if the electrolyte concentrates in the battery, the positive grids are likely to corrode more rapidly, especially during overcharge periods, and again the total life of the battery will be impaired. Since sulfuric acid is not lost from the battery as a result of the direct electrolytic reactions, only water must be replaced periodically to maintain the electrolyte composition and level. During overcharge, gases are evolved as a result of the decomposition of water, and this not only represents a loss of water, but also the gases produce an electrolyte spray above the upper portions of the cell. This electrolyte spray is extremely fine, and so quickly finds its way out of the cell as entrainment with the gases being evolved.

If this should occur, the top of the battery is very likely to get wet with sulfuric acid, which, being a very good electric conductor, permits stray currents to be set up, not only from cell to cell, but also from each cell to ground. With poor top-of-battery cleanliness, these currents can indeed reach very significant values, and the battery can be rather quickly discharged or even destroyed. To avoid this, many manufacturers have developed ingenious spray traps and have incorporated these in their vent plugs. However, careful and proper charging can go a long way toward eliminating this difficulty. Where electrolyte has come out of the cell, it is necessary to neutralize it (boric acid or bicarbonate of soda being recommended) followed by a thorough washing and drying of the top of the battery. This procedure will minimize the destructive effects of corrosion on external battery parts, terminals, and holders.

The loss of sulfuric acid as a result of spray entrainment will, of course, also impair the capacity of the battery. However, this sulfuric acid can be replaced, using chemically pure sulfuric acid, to bring the concentration (or specific gravity) of the electrolyte up to the manufacturer's recommended value.

In doing this, it is best to remove as much of the free electrolyte to which the additional acid is to be mixed to an external vessel. The new acid mixture is properly adjusted for concentration and can then be returned to the battery. If fresh acid is simply added to the battery, the dense, fresh acid will simply settle to the bottom and will probably attack the separators on the way down. The addition of fresh sulfuric acid also generates quite a bit of heat, and, of course, concentrated sulfuric acid is quite dangerous to handle. This operation should, therefore, be conducted only with proper protective clothing and facilities, and under qualified supervision.

**SEALED UNITS** To minimize maintenance, a number of brands of lead-acid batteries have been put on the market as "sealed" or "closed" units. Such batteries are generally made using the calcium alloy system which has an extremely high back emf (of the order of 2.8 to 3.0 V/cell when approaching the full charge condition). For this reason inexpensive automatic charge control equipment can be used to limit the amount of overcharge. This, coupled with the fact that local-action shelf losses are extremely small for calcium alloy batteries, has resulted in a condition in which not a great deal of water is used up during the normal use of the battery, nor is there any serious gassing on standby. Under these conditions the cell can essentially be "closed," since the need to replace water will have been reduced to the point where the cell can go without rewatering for its useful life. The fact that only very little gas is produced and even this will be principally hydrogen (a gas that is very difficult to contain) permits the gas to slowly but surely leak out via even very minute flaws in the seal. Since such batteries can, therefore, go for very extended periods without any maintenance, their use, particularly in small power applications and automobiles, is increasing.

Since many small power applications require a degree of portability (as, for example, power tools, etc.) it is also desirable to immobilize the electrolyte. This is usually done by forming a "gel" in the electrolyte by the addition of silica gel (finely divided  $\text{SiO}_2$ ) to the electrolyte. This in effect absorbs the electrolyte, permitting the battery to be turned even upside down without reorientation of the electrolyte.

However, in immobilized electrolyte-sealed types, it is very desirable to avoid dis-



ruption of the gel, such as would occur with excessive gassing. Consequently such batteries must be recharged with perhaps a little more care than those having free electrolyte, although the usual rules still apply.

**CHARGING NICKEL-CADMIUM BATTERIES** The ability to be recharged quickly, and here we are talking of within an hour or so from a full discharge, is also true for the sealed version. However, in the sealed version, the danger of disrupting the battery physically during or even approaching the overcharge period is greater than in the vented type. The sealed nickel-cadmium battery is made possible by the fact that the negative electrode is capable of recombining the oxygen liberated by the positive electrode during the overcharge period, while at the same time it is possible to arrange the relative capacities of the electrodes so that the liberation of hydrogen at the negative electrode during the overcharge period is entirely avoided.

In addition, local-action losses during storage are such as to avoid the production of hydrogen, still further enhancing the conditions that permit the sealing of such cells. Thus, while a sealed cell can also be recharged rapidly, greater precautions (such as exercise of greater electric control) must be taken to reduce the rate as the cell approaches gassing conditions. Generally speaking, a sealed nickel-cadmium cell will be operated at somewhere in the vicinity of 20 to 30 percent lower available capacity than a similarly sized vented cell. However, the convenience of no physical maintenance, and the cleanliness of external components (i.e., freedom from spray and spillage) have made the sealed nickel-cadmium cell very popular in small power devices where extended cycle life is particularly important.

Although nickel-cadmium batteries are indeed rugged, there is an electrical condition that can quickly lead to the demise of either vented or sealed types. This, known as "runaway," occurs when the battery is charged in such a manner as to create a great deal of internal heat. This can result from excessive charge rates, or from having the battery in an environment where the heat of the charge (particularly during the overcharge period) cannot be readily dissipated. As the temperature of the battery rises, its internal resistance drops very sharply. When the charge is being carried out from a "system-controlled" power source (i.e., essentially a constant-potential charge), a lowering of the internal resistance will result in an increase in the charging current. This increased current in turn results in a lowering of the efficiency of the charge acceptance, still further increasing the heat generation, still further raising the temperature of the cell. The cell can thus very quickly be brought to a condition where the heat becomes physically disruptive. This runaway condition is prevented by providing for better electric controls and adequate ventilation of the cells and the compartment in which it is housed.

An interesting characteristic of nickel-cadmium batteries is the so-called "memory effect" in which the battery seems to "remember" the manner in which it has been cycled. The memory effect shows up only in a highly repetitive cycling routine, and manifests itself as an apparently permanent loss of reserve capacity. Thus, if a 10-Ah battery is cycled on a routine procedure to a depth of only 4 Ah, with the intention to keep the remaining 6 Ah in reserve against emergencies or to enhance long cycle life, the actual available capacity of the battery falls to 4 Ah after a fairly short time on this routine. To avoid this, batteries in such service are periodically given conditioning cycles in which the depth of charge and discharge, and/or the rates at which these are performed, are varied from the routine. This treatment completely avoids the memory effect, and in fact can restore the capacity of a battery so affected, although the latter is more difficult and requires a greater number and variety of conditioning cycles.

**CHARGING SILVER-ZINC BATTERIES** All the general rules for charging secondary batteries apply to silver-zinc types as well, only perhaps more so. Since such batteries are frequently used in higher-voltage systems, in fact up to 150 to 200 V, top cleanliness to avoid leakage currents and stray currents to ground becomes most important. The deleterious effects of excessive overcharge, and particularly high end-of-charge rates, are also magnified in this kind of battery. In overcharge resulting in gassing, the loss of electrolyte can be quite serious since (unlike lead-acid batteries, for example) there is very little free electrolyte (i.e., that which is not absorbed by the separators and electrodes). Any loss of free electrolyte causes drastic changes in the electrolyte level, and

its replacement is difficult, especially if the distribution of the electrolyte among the various electrochemical components has been disturbed. Then too, since the zinc electrode is partially soluble in the electrolyte, a loss of the latter also actually means a loss of electrode material.

A still more serious effect of overcharging can occur where heat is also allowed to build up inside the cell. This is rather easy to do since silver-zinc batteries are almost always housed in plastic cell containers which are poor thermal conductors. A number of close-packed cells, charged and overcharged at excessive rates, will rapidly result in high temperatures, particularly at the center of the battery. If we remember that silver-zinc cells have organic separators and diaphragms under contact pressure with a strong oxidant such as the silver oxide of the positive electrode (see Fig. 17.15), there is little doubt that high temperatures result in the failure of these components.

In charging silver-zinc secondaries, therefore, sufficient time should be allowed between the end of discharge and the start of charge to allow the battery to cool down. A similar rest would also be very desirable between the end of charge and subsequent discharge. Since very little heat gets out around the sides of the battery, a good practice is to blow air gently over the terminals and intercell connectors. These, being of silver or silver-plated copper, can conduct a good deal of heat from inside the battery to where it can be conveniently dissipated.

While voltage-actuated control systems can be used, Fig. 17.16 indicates that the largest voltage rise as a silver-zinc cell is charged occurs when the battery is only about two-thirds charged, and that the rise at full charge is not very pronounced. Also, voltage is not a clear-cut means of establishing the advent of overcharge. Although not shown in Fig. 17.16, the battery voltage at the end of charge is not at all fixed, slowly increasing from about 1.88 V for a fresh battery to even as high as 2.1 V for one that is nearing end of life. For these reasons the most common practice is to recharge silver-zinc batteries at constant current for a fixed period of time, and to do so for such a time as to return 105 to 125 percent of the ampere-hours removed during the previous discharge. A slightly more sophisticated system takes advantage of the fact that silver-zinc batteries will accept a charge at very high rates at the start of a charge (after a deep discharge), and in fact will do so up to the first sharp voltage rise at about two-thirds full charge. This charging system simply uses a high constant-current charge period with a time control to switch to a lower constant-current level for completion of the charge. This two-step system has a great deal of merit, and many manufacturers now recommend such a charging procedure.

## 17.5 SPECIAL TYPES

A considerable amount of research and development is going on in the battery field today relating to new electrode combinations. Since practically any redox reaction can be conducted electrolytically in a cell from which electric energy can be withdrawn, a great many couples (electrode combinations) have been proposed. A few of these have found specific applications. Among the more important are:

1. Magnesium dry cells
2. Organic depolarized cells
3. Cuprous chloride-magnesium cells
4. Silver chloride-magnesium cells
5. Air-depolarized cells
6. Air-zinc cells

The general properties of these batteries are outlined in Table 17.13.

Each of these has at least one and sometimes even several special properties that has been a key factor in its growth. Magnesium dry cells are capable of storage under high-temperature high-humidity conditions; organic depolarized dry cells are potentially capable of very high capacities with light weights; cuprous chloride-magnesium types are capable of operation in very low pressure and low temperature ambients; silver chloride-magnesium batteries are quickly activated by dilute saline solutions; air-depolarized cells require practically no maintenance; air-zinc cells are capable of high energy yields per unit weight; and so on. It is important, therefore, to consider these properties in greater detail, and also in combination with other properties.

TABLE 17.13 Characteristics of Special Cell Systems

Type	Open-circuit voltage, V	Average discharge voltage, V	Capacity, Wh	
			Per lb	Per in <sup>3</sup>
Magnesium dry	1.95	1.60	45	3
Organic depolarizer (magnesium anode)	1.65	1.20	64	3
Cuprous chloride-magnesium	1.5	1.15	31	2
Silver chloride-magnesium	1.9	1.5	75	8
Air-depolarized	1.45	1.05	71	2
Air-zinc	1.45	1.2	52	6

**MAGNESIUM BATTERIES** At the head of the list, in terms of commercial production, is the magnesium dry cell which we have already introduced under primary batteries. This particular battery very much resembles a conventional LeClanche dry cell in which a magnesium can has simply been substituted for the conventional zinc can. This accomplishes several things including raising the voltage of the cell, increasing its power capability, and (as a result of other design modifications) also its capacity.

However, it is the nature of the magnesium anode that it very quickly forms a very protective film when in contact with the battery electrolyte, and it is this film that gives the magnesium dry cell its outstanding property. This film protects the underlying metal and permits the battery to have an extremely long shelf life. It also helps it to withstand very adverse storage conditions such as high temperatures and high humidities. Not a great deal of reliable data are yet pertinent to the shelf life of magnesium dry cells, but there are the indications that they can easily be stored for as long as five to seven years with only modest losses (i.e., 10 to 20 percent for cells of current designs). In terms of adverse storage conditions, magnesium dry batteries can be stored continuously for 30 days (and many for 60 and even 90 days) in ambients of 160°F and 90 to 95 percent humidity with losses of only about 10 percent in capacity; and even this is largely due to loss of moisture instead of anode deterioration. This latter feature makes the magnesium dry cell extremely attractive for use in those areas of the world which are subject to hot, humid climates, and where distribution logistics require a fairly long time interval between manufacturer and user.

**ORGANIC DEPOLARIZED BATTERIES** The organic depolarized cell is physically similar to the magnesium dry cell and, in fact, may even employ the same magnesium anode can, but in this case reducible organic material takes the place of the usual  $\text{MnO}_2$  depolarizer used in magnesium or LeClanche dry cells. Organic depolarizers, in addition to being synthetically produced materials (and, therefore, of more uniform and continuing supply than conventional depolarizers), have the outstanding characteristic of being able to absorb a large number of electrons during the course of their reduction. This produces batteries of very high capacity for a given space and weight. However, because the organic depolarizers are completely nonconductive electrically, the battery must have a substantial grid structure (usually a carbon "chain" as in other dry cells), and this tends to reduce its space advantage. Even so, organic depolarizer cells show a fairly high internal resistance, but where the drain rates are low, as in radio receivers, it warrants, and is receiving, a great deal of attention.

**CUPROUS CHLORIDE-MAGNESIUM BATTERIES** The cuprous chloride-magnesium ( $\text{CuCl-Mg}$ ) cell is again a very specialized type which has found considerable use in radiosonde work. Since it is stored in a dry condition, the  $\text{CuCl-Mg}$  battery is classified as a reserve type, activated by immersion in water (either fresh or saline). For radiosonde work, the  $\text{CuCl-Mg}$  battery uses a separator which is highly absorbent so that the electrolyte remains within the cell after it is withdrawn from the activating liquid. In this condition the cell can be sent to very high altitudes where pressures and temperatures are extremely low. The special characteristics of the cuprous chloride-magnesium battery which permits this to be done is the fact that as soon as



**Fig. 17.18** Cuprous chloride-magnesium radiosonde battery. (Courtesy Clevite Corp.)

the battery starts to discharge, copper is transferred from the cathode to the anode where it catalyzes the local action reactions. The resulting high local action rate produces a great deal of heat which keeps the battery at a suitable operating temperature, in spite of the fact that it may be operating in extremely low ambients.

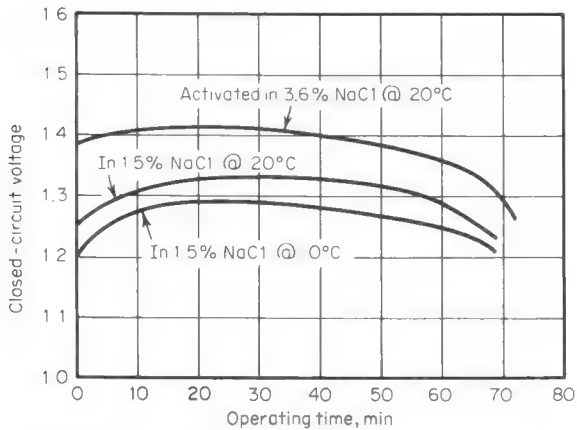
Figure 17.18 shows a typical cuprous chloride-magnesium battery that powers a radiosonde for high-altitude weather observations. This unit supplies both A and B power, the latter being provided by 78 series-connected cells in three parallel stacks. The peculiar spatial arrangement provides the proper balance between the heat generated and the heat lost to the outside necessary to keep the battery at a proper operating temperature.

**SILVER CHLORIDE-MAGNESIUM BATTERIES** The silver chloride-magnesium ( $\text{AgCl-Mg}$ ) battery is, in a general way, similar to the cuprous chloride-magnesium battery, and in fact has been used for radiosonde work where higher power requirements are needed than can be supplied by a  $\text{CuCl-Mg}$  battery. The principal application for silver chloride-magnesium batteries, however, is to power sonobuoys and torpedoes. The  $\text{AgCl-Mg}$  system has an outstanding characteristic in being extremely stable in storage without any special preparation or care, and units that have been stored for more than ten years have given full performance when activated and discharged.

In sonobuoy or torpedo batteries, the saline water of the ocean itself forms the electrolyte. Because this type (dilute salt) of electrolyte is a fairly poor conductor, the electrodes must be spaced very closely together, thus achieving the low internal resistance which makes possible the high power levels at which this type of battery is generally used. Here again local action at the magnesium anode produces a good deal of heat, so that means are usually provided for continuously flushing the battery with sea water. This not only keeps the internal water temperature down, but also keeps the electrodes free of sludge resulting from the discharge. The silver chloride-magnesium battery, the characteristics of which are given in Fig. 17.19, is capable of an extremely wide range of rates, performing well even at very high power levels.

**AIR-DEPOLARIZED BATTERIES** Since the air-depolarized cell actually uses atmospheric oxygen as its depolarizer, it can have very high capacities within a given weight. However, it does not have very high or even medium rate capabilities, and although it has been on the market for a number of years, it has found only limited appli-

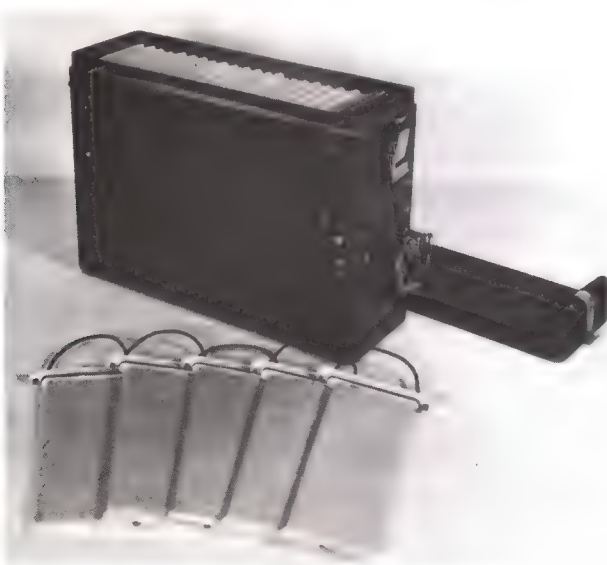




**Fig. 17.19** Characteristics of (sonobuoy type) silver chloride-magnesium cell.

cation in low-power devices at inaccessible locations. The air-depolarized cell uses a cast block of zinc as its anode, and a porous carbon electrode open to the air as its cathode. The electrolyte is potassium hydroxide, and is usually present in large excess of operating requirements so as to practically eliminate watering and other maintenance procedures.

**AIR-ZINC BATTERIES** Recently more sophisticated versions of the air-depolarized cell have been under intensive development, and such cells (renamed air-zinc cells to differentiate them from the earlier versions) are now made in a variety of sizes. These are beginning to find one application in powering radio transceivers. Still further applications are envisioned in the form of motive power (and even automobile) systems, since the cells receive their depolarizer from the air as needed.



**Fig. 17.20** Air-zinc battery with replaceable anodes. (Courtesy Clevite Corp.)

In the modern air-zinc cell, special low-cost cathode catalysts are employed to achieve respectably high rates of discharge. However, no really standardized sizes have yet been presented for general commercial use, since most batteries developed to date have been against highly specific military uses. In a broad sense, air-zinc cells can also be considered to be reserve batteries since they are packaged in dry form and are activated by addition of water. They carry solid potassium hydroxide (usually in the pores of the anode), which by combining with the water forms the electrolyte of the cell. After activation, the shelf life of the battery is fairly poor, and pretty nearly all its capacity will be dissipated in a matter of weeks. However, some considerable work is now being done to improve this situation, principally through amalgamation of the zinc with mercury, a process that greatly reduces the rate at which the zinc will spontaneously react with the electrolyte.

In addition to improvement of the primary version, work is also being directed toward recharging the zinc electrodes in air-zinc cells in a more or less conventional manner, thus producing a secondary battery. At the present time such batteries have not reached any really commercial performance, but an alternate development, the so-called mechanically rechargeable battery, has become available. In this type of battery, illustrated in Fig. 17.20, the zinc anodes are designed in such a way as to be easily removed from the battery upon completion of their discharge, and new ones are slipped into their place. During this process the electrolyte is also replaced, and the result is a battery virtually capable of brand-new performance. The positive, or air electrode, has been developed to a current life expectancy equivalent to about 50 anodes, making this a fairly economical battery to operate.

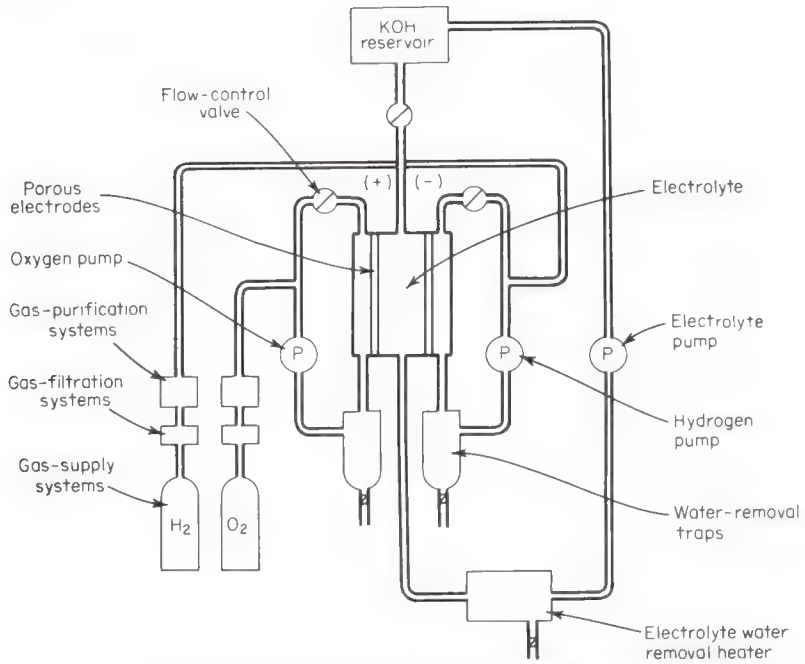
## 17.6 FUEL CELLS

The discussion of air-zinc cells, where one of the reactants (oxygen) enters the cell as needed, leads directly into a group of electrochemical generators usually referred to as fuel cells. As indicated earlier, many electrochemical reactions are reversible. From this the concept was evolved that if water can be broken down electrochemically into hydrogen and oxygen, it should be possible to recombine hydrogen and oxygen in an electrochemical cell (i.e., on suitable electrode surfaces) so as to yield electric energy. This indeed proved to be true, and cells were developed as early as the 1920s which accomplished the pertinent reactions at reasonably good efficiencies. In a great many respects, such cells had the characteristics of a prime mover-generator combination since the electrochemical materials (hydrogen and oxygen) were gases, and could be stored outside the cell to be metered into the cell as needed. Within limits it was found that the greater the rate at which the gases were fed, the greater was the electric power generated by the cell. Furthermore these gases acted very much as they might have as a fuel and an oxidant, and hence the popular term fuel cell.

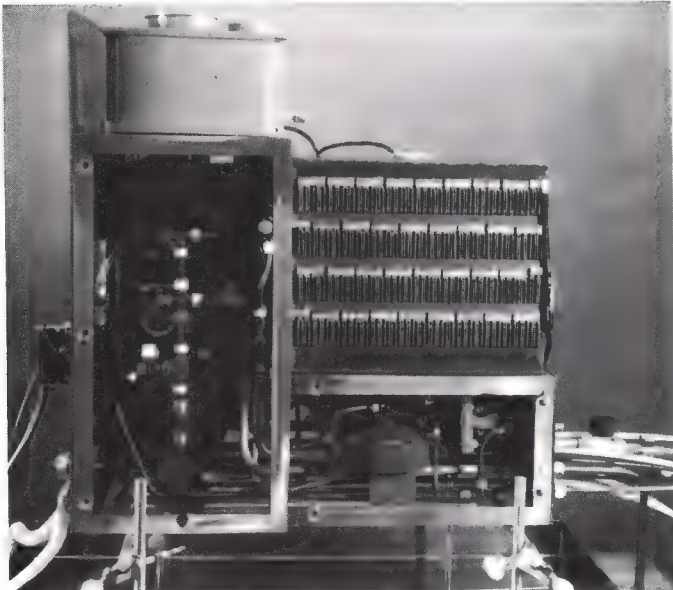
In more recent times, a great variety of "fuels" have been found to be suitable, as have a fairly large number of oxidants, so that by now there is a variety of fuel-cell systems. Also, the concept has been enlarged to embrace other than gaseous fuels, and typically, in addition to hydrogen, such materials as hydrocarbons, hydrazine ( $N_2H_4$ ), and molten metals (such as sodium, lithium, potassium, and other reactive easily fused metals) have been used. Oxidants such as air, chlorine, bromine, and others, have also provided interesting cell characteristics. As a matter of fact, the various fuel-cell combinations of oxidants and reductants that have been investigated have resulted in fuel-cell systems with highly specialized characteristics for very specific applications.

Basically, however, pretty nearly all fuel cells are variations of the fundamental design shown schematically in Fig. 17.21. This figure shows a hydrogen-oxygen cell and the auxiliary equipment needed to support it. This auxiliary equipment turns out to be extremely important for the extended operation of a fuel cell. In the cell illustrated, the overall reaction involves the oxidation of hydrogen to water. This water enters the electrolyte and, thereby, dilutes the latter. Some system must be provided, therefore, to remove the water in order to maintain the electrolyte concentration at appropriate levels. If the hydrogen or oxygen contains any impurities, these will enter the cell and ultimately build up to the point where they will seriously impair the performance of the cell, and the fuel and oxidant must therefore be brought to a high level of purity before entering the cell. These turn out to be very difficult engineering problems, and a good





**Fig. 17.21** Schematic diagram of hydrogen-oxygen fuel-cell system.



**Fig. 17.22** Hydrogen-oxygen fuel-cell power system. (Courtesy Union Carbide Corp.)

TABLE 17.14 Characteristics of Typical Fuel-Cell Systems

Power capability, W	Application	Fuel	Oxidant	Unattended life, h	Power capability	
					Kw/lb	Kw/R <sup>3</sup>
200	Radio communication	Hydrogen	Oxygen	1500	0.006	0.4
500	General power	Hydrazine	Air	1000	0.005	0.2
10 000	Space power	Alcohol	Air	2500	0.000	0.3
25 000	Marine	Hydrocarbon	Air	7500	0.01	1.3

deal of the progress that has been made has resulted from careful development of the auxiliary equipment. A photograph of an actual fuel-cell system is shown in Fig. 17.22, with general operating characteristics of a number of fuel-cell systems detailed in Table 17.14.

Fuel cells are, therefore, no longer simple electrochemical power generators in that the movement of fuel and oxidant into the cell and the products of the oxidation out of the cell have become a fairly complicated materials movement problem, and we find that modern fuel cells have a great deal of pumping, filtering, throttling, and control equipment associated with the electrochemical generator itself. In many instances, as for the  $H_2O_2$  cell, it has been necessary to go to very high pressures and temperatures (to achieve high power outputs) further complicating the equipment. Others, such as the lithium-fluorine system, require extremely high temperatures in order to operate properly.

Many fuel cells are now available with fairly large power capabilities and have found use as the power source for manned space vehicles, as for example in the Apollo program. Fuel cells using alcohol have been produced in smaller sizes and have been promoted for military communication. To date, fuel-cell designs are developed against very specific purposes, and there are no generalized sizes, or standardization of components, so that tabularized information of this nature is not currently available. Nevertheless the wide variety of properties that can be achieved, coupled with long-life catalysts which greatly increase the rate of reaction possible at the electrode surfaces (and thereby increase the power capability of the cell), have accounted for the tremendous current interest being shown in these devices. In addition, a great many organizations, both government and private, are supporting major projects to increase still further the efficiency and the ease of the operation of these units, and it is very likely that standard forms will be evolved over the next few years.

At the moment, the indications are that these forms will be directed particularly toward motive power applications (small automobiles and the like), stationary generators (particularly for standby service), and radio communication. Fuel cells have indeed shown themselves to be capable of producing usable quantities of energy directly from conventional fuel sources and have a long life with relatively little maintenance. As a matter of fact, it is now entirely feasible to have completely unattended remote fuel-cell power plants, or to utilize waste fuels (such as hydrogen from caustic production) to produce "convenient" power. At the present time, however, the cost of fuel cells and their operation are high, but there is every expectation that as their use becomes more widespread and improvements continue, and as some degree of standardization is achieved, fuel cells will occupy an important position as electrochemical power sources.



# Chapter 18

## Microprocessors and Microcomputers

**Gary C. Meyer**

*Central Region Service Manager  
Philips Test & Measuring Instruments, Inc.  
Minnetonka, Minnesota*

### 18.1 INTRODUCTION

Our world is exploding with technology that has made computer a household word. Semiconductor evolution has advanced from the discrete transistor to the integrated circuit (IC). The heart of modern computers is an IC known as a microprocessor. The microprocessor (sometimes referred to as a central processing unit, or CPU) is also being used in many different applications. This device is used in "on-board" automotive computers, sewing machines, process control instrumentation, burglar alarms, oscilloscopes, programmable temperature controllers, personal computers, video games, and numerous other electronic systems.

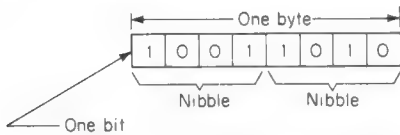
Microprocessors integrate registers, limited memory, an address bus, a data bus, a control bus, timing control circuits, and arithmetic and logic circuits all in a single chip. In addition to the microprocessor, a computer requires peripheral devices plus additional memory to store programs and data. Examples of peripheral devices include digital test instruments, printers, cathode-ray tube (CRT) displays, floppy disks, and cassette tape.

Several manufacturers produce microprocessors that offer a diversity of operational techniques and specifications. Leading manufacturers include Intel Corporation, Motorola, Rockwell, and Zilog. Documentation available from manufacturers includes some types of programming manual and specification sheets for each individual device.

This chapter discusses the Intel 8085, as a model, including hardware and software (the written instructions) to provide a basis for the study of other microprocessors. The chapter concludes with an examination of such popular CPUs as the Zilog Z80, Motorola 6800, and the Syntek 6502.

### 18.2 TRANSPORTING DIGITAL WORDS

A microprocessor is a relatively unintelligent device that requires precise instructions to function properly. The only language it understands is "machine language" consisting of 0s and 1s (binary numbers). Combinations of 0s and 1s comprise a digital word. Refer to Fig. 18.1. Each box represents a *bit*. And eight bits is referred to as a *byte*, and 4 bits is called a

**Fig. 18.1** An example of an 8-bit (1-byte) word.**Fig. 18.2** Dividing the word in Fig. 18.1 into two 4-bit fields.

**nibble.** Digital words are simply a group of bits treated as a unit. A 16-bit word, for example, is 2 bytes wide.

The digital word in Fig. 18.1 represents a decimal number which may be determined as follows:

$$\begin{aligned}
 10011010_2 &= (1 \times 2^7) + (0 \times 2^6) + (0 \times 2^5) + (1 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) \\
 &\quad + (1 \times 2^1) + (0 \times 2^0) \\
 &= 128 + 16 + 8 + 2 = 154_{10}
 \end{aligned}$$

where subscript 2 indicates a binary number and subscript 10 a decimal number. This binary number (or decimal number) can be converted to hexadecimal number 9AH.

Digital words used in microprocessor applications are divided into groups of four and referred to as a *4-bit field*. Each 4-bit field also can be represented by binary-coded decimal (BCD) notation. BCD should not be confused with binary numbers. For example, dividing the digital word of Fig. 18.1 into two 4-bit fields is illustrated in Fig. 18.2. Each field can be reduced to hexadecimal notation by finding the equivalent BCD in Table 18.1 and writing out the corresponding number in hexadecimal notation.

Consider the first field of our example. A BCD of 1001 appears opposite number 9 in Table 18.1. The second field of 1010 is opposite number A; hence, the hexadecimal is 9AH, or simply 9A if hexadecimal notation is understood.

Referring to Fig. 18.3, we see the voltmeter can be switched to any line, thus monitoring whether the line is high or low. As switches S1 through S4 are changed, decimal numbers 1 through 15 or hexadecimal numbers 0 through F can be represented. The lines are also identified by the least significant bit (LSB) and most significant bit (MSB). This identification does have value in relating the BCD notation to the hardware of Fig. 18.3. LSB refers to the far right bit of the BCD, while MSB refers to the far left bit.

**example 18.1** For the circuit of Fig. 18.3, which switches must be closed and which open to produce the 4-bit field 1010? What are the LSB and MSB values? What hexadecimal number is represented?

**solution** In general, the data field takes on the format

**TABLE 18.1** Decimal, BCD, and Hexadecimal Equivalents

Decimal	BCD	Hexadecimal
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

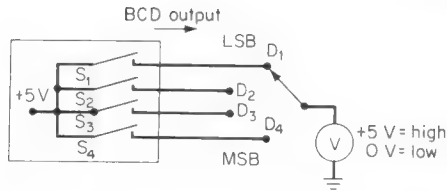
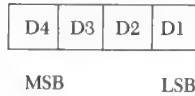


Fig. 18.3 Relationship of BCD to hardware.



For 1010,  $D4 = 1$ ,  $D3 = 0$ ,  $D2 = 1$ ,  $D1 = 0$ . Therefore,  $S4 = \text{closed}$ ,  $S3 = \text{open}$ ,  $S2 = \text{closed}$ ,  $S1 = \text{open}$ . Also  $LSB = 0$ ,  $MSB = 1$ , and hexadecimal number = A.

The connecting wires of Fig. 18.3 are called a *bus*, or *data bus*. In this example, it is called a four-wire bus. Since data is flowing in only one direction, the term *unidirectional*, or one-way, bus is applied. Obviously eight wires can be used to represent two 4-bit data fields. This is the case in many microprocessor systems.

If several devices are to be tied together in a system, the same bus may be used to communicate between devices. It is also convenient to show a bus as two parallel lines instead of drawing 4, 8, or 16 wires on a diagram. Refer to Fig. 18.4. Device 1 is using a one-way, or unidirectional, bus to the main data bus. Devices 2 and 3 are shown with double arrows, indicating they can output, or receive, data from the main bus. This is called a *bidirectional*, or two-way, bus. Devices 2 and 3 can receive data from device 1 or communicate with each other.

Note that only two circuits can "talk" to each other on the bus at a given time; therefore, precise timing and control functions are accomplished by adding two additional buses (control and address), as shown in Fig. 18.5.

Each device is assigned a fixed number, specified as an address, just as houses on a city block can be identified by a house number and street. The timing and control circuit controls which devices are connected to the data bus first by giving the address, second by controlling the direction of data flow, and third by providing exact timing via the control bus.

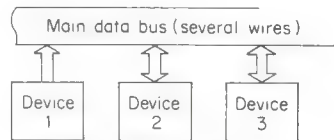


Fig. 18.4 Example of unidirectional and bidirectional buses.

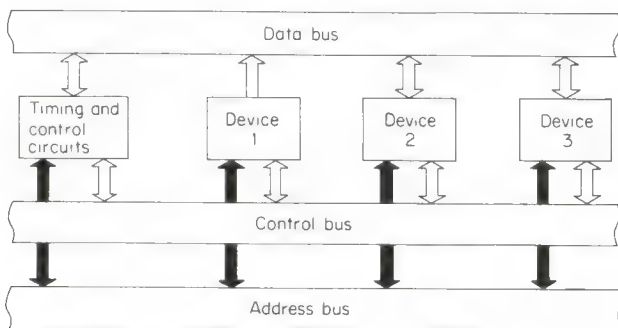


Fig. 18.5 A total bus system.



### 18.3 INTRODUCTION TO MICROPROCESSOR SYSTEMS

Before we examine a microprocessor system, a checkbook analogy is presented. An individual keeping a checkbook current is required to enter the amount of a check and subtract that amount from the previous balance. Referring to Fig. 18.6, assume the individual (operator) has never used a checkbook or a calculator. Very detailed instructions, therefore, must be given. The operator can only read instructions, perform instructions, read numbers, and write down numbers. All instructions must be carried out in the exact sequence specified to achieve correct results. Note that every instruction is identified by a fixed number. Specific line numbers (or check numbers) in the checkbook also identify a location on the page. Microprocessor systems work in much the same way; Table 18.2 summarizes the comparison.

A CPU requires a detailed set of instructions called a *program*. The program sequence is controlled by assigning a number called an *address* to each line. Programs for a CPU can be written in assembly language by using shorthand words called *mnemonics*. Programming is covered later in this chapter.

A block diagram of a typical microprocessor system is shown in Fig. 18.7. The read-only memory (ROM) and/or part of the random-access memory (RAM) contains the set of instructions to run the system in a given sequence. Each location in these memories is given a specific address with data at each address. The CPU initiates the program in the following manner:

1. The CPU outputs the starting memory address on the address bus.
2. The starting address located in memory is recognized and outputs its corresponding data onto the data bus.

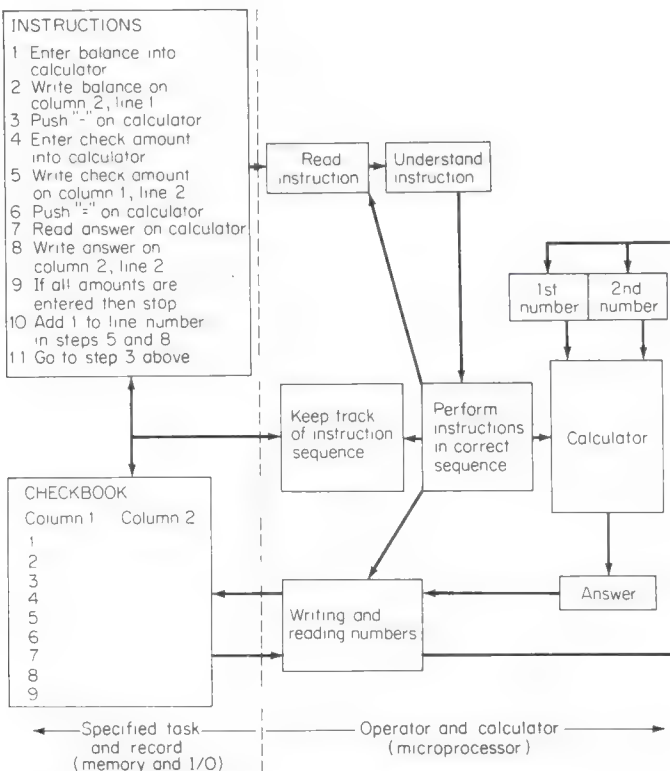


Fig. 18.6 Checkbook-microprocessor analogy.

**TABLE 18.2 Summary of Checkbook-Microprocessor Analogy**

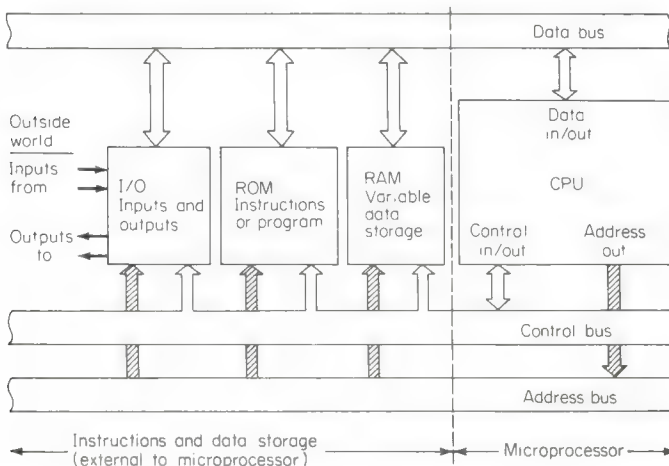
Real-life terms (see Fig. 18.6)	Microprocessor system terms
Set of instructions	Program or software—stored in memory
Instruction line numbers	Program addresses
Checkbook	Data stored in RAM memory
Check numbers	Address where data is stored
Read instruction	Instruction register
Understand instruction	Instruction decoder
Perform correct subsequence	Timing and control
Keep track of sequence	Program counter
Writing and reading	Write data to/read data from RAM
First number/second number/ answer	Input-output ports
Calculator	Temporary registers Arithmetic logic unit (ALU)

3. The CPU inputs the data from the data bus, interprets the data, performs an operation, and places the next address on the address bus.

During this procedure, the control bus receives information from the CPU to enable the corresponding external block [that is, input-output (I/O), ROM, or RAM] at the appropriate time. Other control lines output signals to I/O or memory to enable a read or write function. *Read* means that data is being received from the memory by the CPU; *write* means that the CPU is sending data to memory. All operations must take place at a specified time determined automatically by the CPU. Since the processor performs operations very quickly (usually in a few microseconds), sometimes the control lines are used to delay the CPU until slower peripheral devices complete their operations.

An I/O device shown in Fig. 18.7 extends the system capability to input data from, or output data to, the outside world. This term refers to peripheral devices such as CRT displays, control circuits, tape recorders, and printers. Also I/O ports have addresses to identify their location.

**MICROCOMPUTERS** The microprocessor system shown in Fig. 18.7 is the heart of a microcomputer. Some type of keyboard is connected to one I/O input port for programming, and a CRT readout is interfaced with another I/O output port. A printer also may be driven from another I/O output port to produce *hard copy*, a name given to the printed page. Microcomputer ROMs contain programs to read the keyboard and control the display.

**Fig. 18.7** Block diagram of a basic microprocessor system.

## 18-6 Microprocessors and Microcomputers

Often system RAMs are used to store a program inputted from the keyboard, magnetic tape, or disks which are then used to run the microcomputer.

Computer specifications include the amount of storage available in terms of 1K, 4K, 16K, 32K, or 64K bytes of memory. These notations are rounded-off numbers specifying the approximate number of memory locations (1K byte = 1024).

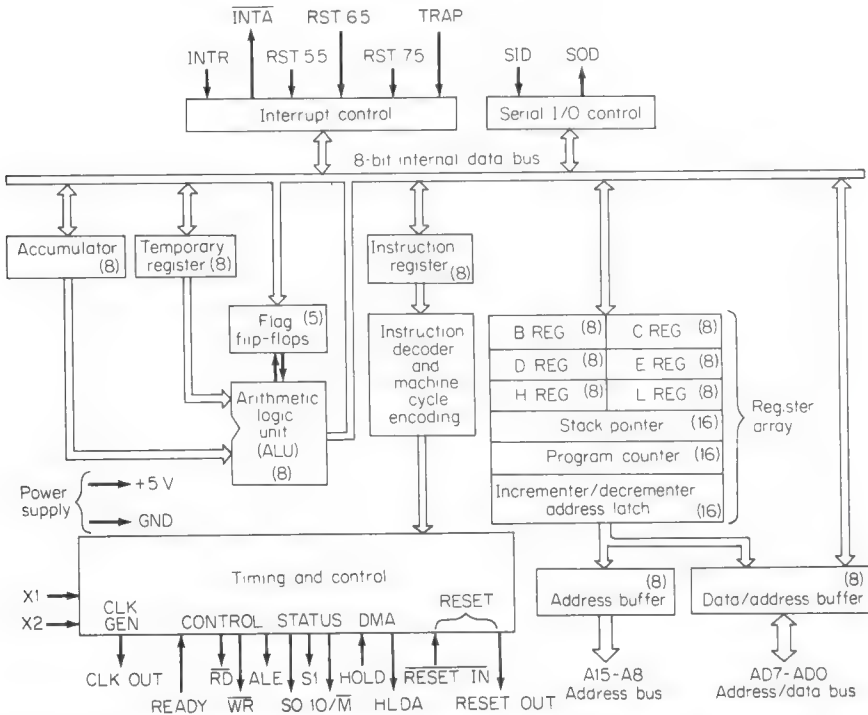
Additional hardware outside the microprocessor system is often required to complete a microcomputer. For example, the CRT must have additional control circuitry and a character generator to produce a display. Keyboards also have associated circuitry to decode the switches during typing. If BASIC computer language is used to program the system, an additional ROM program is supplied to assemble or change the keyboard entries to digital words of 1s and 0s which a microprocessor understands.

In its simplest form, a microcomputer is a microprocessor system interfaced to external input-output peripheral devices for programming and readout.

### 18.4 THE 8085 MICROPROCESSOR

A detailed discussion of the block diagram (Fig. 18.8) will give a better understanding of how the microprocessor is able to control digital words (addresses, data, or control) on the three main bus lines. Obviously, measurements cannot be made inside the chip; however, internal data can be placed on the bus lines and examined with a logic analyzer or a development system. Logic analyzers are covered in Chap. 19.

**REGISTERS** Examination of Fig. 18.8 shows several registers used for temporary storage of data during the execution of a program. The register array consists of six 8-bit registers designated by the letters B, C, D, E, H, and L and three 16-bit registers named stack pointer, program counter, and address latch incrementer-decrementer. All 8-bit registers in the register array can be combined in specific pairs to form 16-bit registers if desired. Registers B



**Fig. 18.8** Functional block diagram of the 8085 microprocessor. (Courtesy Intel Corporation.)

and C, D and E, and H and L are the specified pairs. The lettered registers are sometimes called *scratchpad* registers since they are used to store data for short intervals. The HL register pair can be used as a special memory pointer to specify an address in memory which becomes the source or destination of data.

The *accumulator* is an 8-bit register serving some special functions:

1. All arithmetic and logic operations performed in the ALU are based on data in the temporary register and accumulator. Note that the temporary register stores data captured from the internal data bus.
2. Answers from the ALU are returned to the accumulator.
3. Some special load, store, input, and output instructions use the accumulator (see Sec. 18.5).
4. Data bits in the accumulator can be rotated left or right by software or internally in some arithmetic operations. Rotation shifts data bits one position to left or right as specified.

Another 8-bit register is the *flag flipflops* consisting of 5 useful bits and 3 undefined bits. Each bit is either *set* (logic 1) or *reset* (logic 0), depending on the result of a given operation. Data bits in the flag register can be pictured as shown in Fig. 18.9. The five defined flag bits are used to indicate some operation or status of the accumulator register. Bits that are defined indicate the following:

**S** Sign flag D7 (MSB). The sign bit duplicates the accumulator sign bit and stores the result for future reference. This permits negative or positive numbers to be handled by the accumulator from  $-128$  to  $+127$ . Changes in this bit occur after each execution of arithmetic or logic operations.

**Z** Zero flag D6. If certain operations leave all 0s in the accumulator, the zero flag will be set (logic 1).

**AC** Auxiliary Carry (AC) flag D4. The AC flag is set (logic 1) by an overflow out of bit 3 in the accumulator.

**P** Parity flag D2. If the number of 1 bits in the accumulator is even (even parity), this flag is set. If odd parity occurs, the flag is reset (logic 0).

**CY** Carry flag D0 (LSB). Arithmetic operations affect this flag. An overflow out of accumulator bit D7 (MSB) sets the flag.

Many of the arithmetic and logic operations affect one or more flags. A complete description of this interaction is given in the Intel 8080/8085 Assembly Language Programming Manual.\*

Sometimes the accumulator and flag registers are considered together as a 16-bit register and given the name *program status word* (PSW). Under certain program conditions, it is desirable to save the accumulator and flag register contents on the stack. This is accomplished by storing the PSW (16 bits) or register pair in a portion of RAM called *stack*, discussed in the next section.

**STACK POINTER** Consider Fig. 18.10 where information from the CPU register pairs can be saved on the stack. The stack pointer (SP) keeps track of the address in RAM where the last register information is saved. Each time data is "pushed" onto the stack, the SP register is decremented by 2 (subtracts 2). When data is "popped" off the stack back to the CPU, the SP is always incremented by 2. A programmer must initialize the SP in the program, giving the starting stack address.

**PROGRAM COUNTER** A 16-bit register called the *program counter* (PC) is used to keep track of the program instruction sequence. It points to the memory location containing the next instruction to be executed.

D <sub>7</sub>	D <sub>6</sub>	D <sub>5</sub>	D <sub>4</sub>	D <sub>3</sub>	D <sub>2</sub>	D <sub>1</sub>	D <sub>0</sub>
S	Z	-	AC	-	P	-	CY

Fig. 18.9 Example of a flag register.

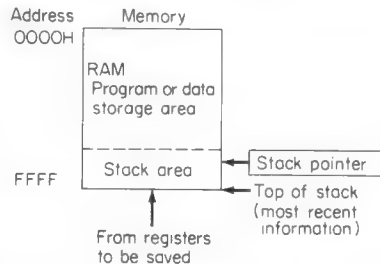


Fig. 18.10 Operation of the stack pointer (SP).

\* Intel 8080/8085 Assembly Language Programming Manual, Intel Corp., Santa Clara, Calif.

**ARITHMETIC LOGIC UNIT** The ALU, as mentioned previously, inputs the accumulator and temporary register, performs an arithmetic or logic operation, and then returns the result to the accumulator.

**ADDRESS LATCH INCREMENTER-DECREMENTER** The address latch outputs an address for a fixed period of time to the address buffers. This 16-bit address may be generated by the program counter, stack pointer, or any 16-bit register. The incrementer-decrementer section provides a way to increase or decrease the register contents by one count at a time.

**DATA OR ADDRESS BUFFER** Addresses from the 16-bit address latch are separated into two bytes called the *high byte* and *low byte*. The high byte (A15-A8) is sent to the 8-bit buffer which is connected to the high-order address bus. The low byte (A7-A0) uses a separate 8-bit buffer which shares data output into the address or data bus at different points in time. This sharing is called *multiplexing* and is discussed later.

**INSTRUCTION REGISTER AND DECODER** The first instruction fetch in a program places a coded instruction on the internal data bus which is transferred to the instruction register. An instruction decoder examines the code and transforms the bits to a language understood by the timing and control circuits. Depending on the instruction, the decoder outputs control the registers, ALU, data buffer, and address buffers.

**TIMING AND CONTROL** Timing signals originate in an internal clock generator, and the clock frequency is determined by an external crystal. Several inputs and outputs accessible to the user provide a means of control or a readout of the CPU status. Figure 18.8 shows the following inputs and outputs:

CLK OUT	Clock out.
READY	Ready input used for external synchronization of slower peripherals.
RD	Read output present when reading from memory or I/O port, active low (as indicated by bar over letters).
$\overline{\text{WR}}$	Write output, active low, present when writing information to memory or I/O port.
ALE	Address latch enable, used to latch the low byte (A7-A0) signal.
S0, S1	Status 0 and status 1 lines indicate machine cycle operations taking place inside the CPU.
IO/ $\overline{\text{M}}$	Input-output or memory status line; a <i>high</i> indicates I/O, a <i>low</i> indicates a memory read or write. IO/ $\overline{\text{M}}$ , S0, and S1 are often used together to indicate other status conditions. (See Fig. 18.13.)
HOLD	An external device is permitted to take over the bus lines; an input signal disconnects the CPU from bus lines.
HLDA	Hold acknowledge indicates the CPU is in hold state.
$\overline{\text{RESET IN}}$	Reset in (active low) sets the program counter to zero and resets interrupt enable and HLDA flipflops. Also it may affect internal register contents.
RESET OUT	Reset out indicates CPU is being reset; it may be used for system reset signal.

**INTERRUPT CONTROL** The interrupt control block inputs provide a means of interrupting normal program flow. Five interrupt codes are given different priorities, as shown in Table 18.3.

**TABLE 18.3 Priorities of Interrupt Codes**

Priority	Symbol	Vectored address
1	TRAP	0024H
2	RST7.5	003CH
3	RST6.5	0034H
4	RST5.5	002CH
5	INTR	Specify



TRAP is a hardware-controlled interrupt. A signal applied to the TRAP interrupt causes the PC to jump to address 0024H. If this interrupt is used in a system, the interrupt service program routine must be written at this address.

Interrupts RST7.5, RST6.5, and RST5.5 are software-controlled. These interrupts are also vectored to specific addresses as listed and are maskable (can be turned on or off by software). One interrupt can also interrupt another, depending on priorities. This is called *nesting*.

The fifth input, INTR, is also enabled or disabled by software. Bytes 2 and 3 of this instruction allow the programmer to specify the interrupt service routine address.

**SERIAL I/O CONTROL** Serial input data (SID) from the outside world is converted to parallel data in this block. The resultant parallel data is fed to the internal data bus. This provides a useful conversion for data transmitted on a two-wire communication system, such as a telephone line, or outputted from a cassette recorder to be used by the microprocessor. Serial output data (SOD) provides a means of feeding the parallel bus data into a two-wire system. Essentially, this block is a communications terminal.

**BASIC SYSTEM TIMING** Each device in a microprocessor system has a unique timing diagram. Clock signals generated by the processor control the exact time that data is placed on, and removed from, the bus. Data on a bus must be valid before it is stored in memory or sent to an I/O port. Similarly, data from memory or an I/O port must be valid on the bus before it is read by the processor. The time required to fetch and execute an instruction is called an *instruction cycle*. Several terms associated with processor timing are shown in Fig. 18.11.

Instruction cycles are subdivided into machine cycles (M). Every machine cycle completes a read or write operation. Read and write operations provide the necessary communication between the processor and other devices to execute any instruction. One to five machine cycles may be required to complete an instruction cycle, depending on the instruction being executed. Figure 18.11 shows an instruction cycle consisting of three machine cycles. The number of states within a machine cycle varies with the operation being carried out.

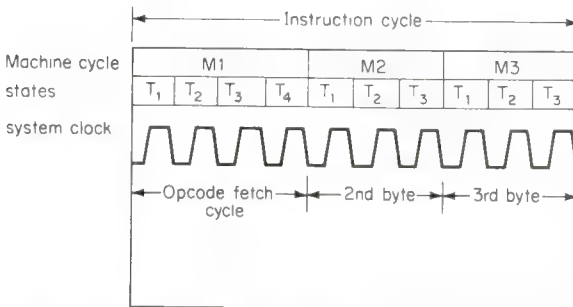
**example 18.2** If the clock frequency is 3 MHz, what is the total time required for the instruction cycle shown in Fig. 18.11?

**solution**

$$T_1 = \frac{1}{3 \times 10^6} = 0.33 \mu\text{s}$$

A machine cycle is  $M = \text{number of states} \times \text{time per state}$ .  
Hence,

$$\begin{aligned} M1 &= 4 \times 0.33 = 1.32 \\ M2 &= 3 \times 0.33 = 0.99 \\ M3 &= 3 \times 0.33 = 0.99 \\ \text{Instruction cycle} &= 3.30 \mu\text{s} \end{aligned}$$



**Fig. 18.11** Example of an instruction cycle.



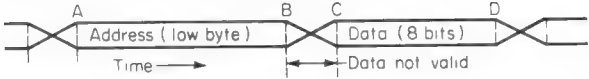


Fig. 18.12 Illustration of multiplexing.

**MULTIPLEXING** Multiplexing is a process of sending more than one set of data over the same bus at the same time. The multiplexed microprocessor uses a time sharing technique made possible by the short time required to execute instructions. As previously discussed, the address/data bus uses this method to place the address low byte on the bus first, followed by data; Fig. 8.12 illustrates the process. Note that for a short time the data is invalid. If the bus is sampled or strobed between A and B, the address (low byte) is seen. If the sampling is performed between C and D, data is extracted. By using the symbology of Fig. 18.12, a timing diagram for the 8085 can be constructed (Fig. 18.13).

The system clock generates timing pulses in which each cycle constitutes one timing state. Processor output A8–A15 is the high-order address which changes with every machine cycle. A0–A7 is the low-order address of the program counter placed on the AD bus. To capture 16 bits (2 bytes) of address information before the AD bus is multiplexed, the address latch receives a signal called the *address latch enable* (ALE). When the ALE signal produces a negative slope (the trailing edge of the pulse), bus lines A8–A15 and AD0–AD7 are sampled and latched. Now data can be placed on the AD bus.

RD is the read signal (active low) which is used to enable the memory when data from the latched address location is present on the bus. The microprocessor must have valid data to feed the instruction register; therefore, it waits for the third state, T3, before sampling the data. Once the data is in a register, the read pulse goes high.

WR is the write signal (active low) which is used to enable the memory to receive data from the microprocessor.

IO/M is a single line used to indicate an input-output operation when the signal level is high or a memory operation when the level is low. In conjunction with the RD and WR signals, the IO/M ensures proper selection of devices and operations.

The bottom line of Fig. 18.13, labeled STATUS, is the control bus consisting of S0, S1, IO/M, RD, WR, and INTA. Table 18.4 shows the logic levels present on the bus for seven possible types of machine cycles.

**example 18.3** The following data bits are observed on the bus lines during M1. (a) What is the 16-bit address? (b) What type of operation is being executed? (c) What is the data expressed in hexadecimal numbers?

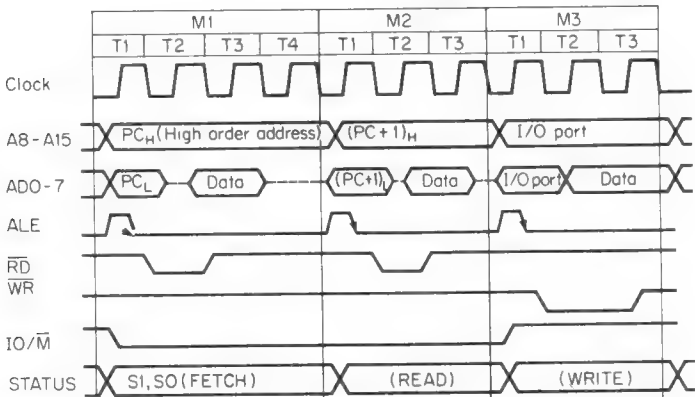
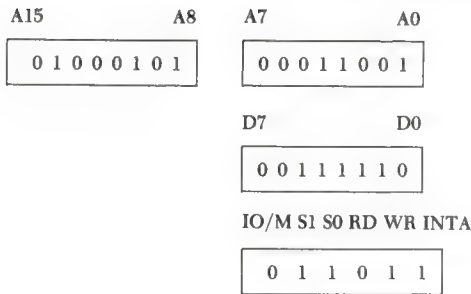
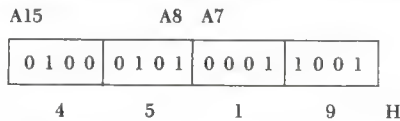


Fig. 18.13 Timing diagram for the 8085 microprocessor.



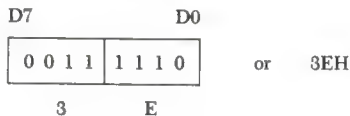
**solution** (a) Bits A15–A8 represent the high-order byte of the program counter (PC<sub>H</sub> of Fig. 18.13). Bits A7–A0 are the low-order byte (PC<sub>L</sub>). The complete address is determined as follows:



First, the bits are divided into 4-bit fields. Next, the 4-bit fields are converted to hexadecimal numbers; hence, the address is 4519H.

(b) Table 18.4 is used to find the operation. Comparison of the data bits shown above with this chart indicates an OPCODE FETCH.

(c) The solution procedure is the same as in (a):



**TYPES OF ADDRESSING** Addressing is the process of pointing to the location (address) of the instruction to be executed or data to be processed. Four basic types of addressing are employed in microprocessors:

1. Immediate addressing
2. Direct addressing
3. Indirect addressing
4. Indexed addressing

**TABLE 18.4 Machine Cycle Chart for the 8085 Chip**

Machine cycle operation		Status			Control		
		IO/ $\overline{M}$	S1	S0	$\overline{RD}$	$\overline{WR}$	$\overline{INTA}$
Op code fetch	(OF)	0*	1†	1	0	1	1
Memory read	(MR)	0	1	0	0	1	1
Memory write	(MW)	0	0	1	1	0	1
I/O read	(IOR)	1	1	0	0	1	1
I/O write	(IOW)	1	0	1	1	0	1
Interrupt acknowledge	(INA)	1	1	1	1	1	0
Bus idle	(BI): DAD	0	1	0	1	1	1
	INA(RST/TRAP)	1	1	1	1	1	1
	HALT	TS†	0	0	TS	TS	1

\*0 = logic 0

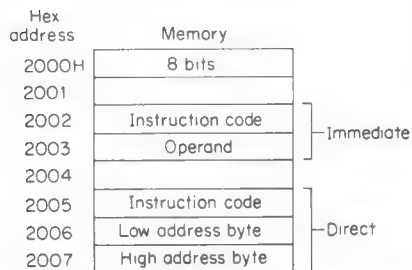
†1 = logic 1

†TS = high impedance

## 18-12 Microprocessors and Microcomputers

The 8085 uses the first three, which are discussed here. Indexed addressing is covered in Sec. 18.8.

Immediate addressing means that the data word, called an *operand*, is located in the memory location immediately following the instruction code. As an example of this type of addressing, assume that the number FF is to be placed in the accumulator during a program. The instruction code (also called operational code, or *op code* for short) for moving data immediately to the accumulator is 3E hexadecimal (00111110). Instruction 3E placed at address 2002 must be followed by the operand or data of FF (11111111) at the next address, 2003. (See Fig. 18.14.) Addressing of this nature is often used to set the contents of a register.



**Fig. 18.14** Immediate and direct addressing.

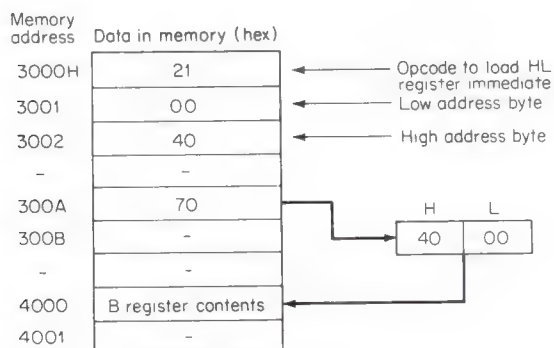
Direct addressing uses a different technique. Immediately following the instruction code in memory is the address of the operand. An example of this technique is the jump instruction. Address 2005 in Fig. 18.14 could be the code for an unconditional jump instruction, or C3 hexadecimal (11000011). Then address 2006 must contain the low byte and 2007 the high byte of the address specified by the programmer. This sequence would cause the program counter to jump to the new specified address.

Indirect addressing utilizes the processor's 16-bit registers to point out an address in memory. First, the desired address is stored in the CPU register. Second, an instruction code, including a reference to the proper register, is stored in memory. If an operand is needed several times in a program, this method saves program length and memory space. For example, consider Fig. 18.15. Moving register B contents to a memory location (M) in an 8085 system can be accomplished as follows:

1. Load the 16-bit register pair HL with the address 4000. The HL register pair is used as a memory pointer.
2. When the processor comes to 300A in the program, it senses the op code for a register B to memory move (70).
3. The processor interrogates the HL register for an address (4000) and duplicates B register contents into that memory location.

**example 18.4** Refer to Fig. 18.15. If the B register contents are to be placed at address 4001 (by using indirect addressing), what address location and what data must be changed?

**solution** Because 4001 instead of 4000 is the new address, the low-order address byte located at memory address 3001 must be changed from 00 to 01. No other change is required.



**Fig. 18.15** Example of indirect addressing.

## 18.5 THE 8085 INSTRUCTION SET

Every microprocessor has an instruction set which defines all the operations possible. Instruction sets are copyrighted by the manufacturer; therefore, similar instructions for various processors will have different names. A thorough understanding of the instruction set is required to write a program.

The instruction set for the 8085 is divided into four groups:

1. *Data transfer group.* Encompasses all the register and memory moves, plus load and store operations.

2. *Arithmetic and logic group.* Addition, subtraction, incrementing, decrementing, rotate, and all logic operations are grouped under this heading. A few specials are also included.

3. *Branch control group.* Includes jumps, calls, returns, and restarts.

4. *I/O and machine control.* Stack operations, I/O, interrupt controls, no operation, and halt operations are covered.

This section lists instructions for the four groups, defines the symbology, and provides some examples to illustrate the required sequence of operation.

**DATA TRANSFER GROUP** This group contains move and transfer instructions used to transfer data internally between registers or externally between registers and memory:

Mnemonic	Meaning
MOV	Move
MVI	Move immediate
LDA	Load accumulator directly from memory
STA	Load accumulator directly in memory
LHLD	Load H and L registers directly from memory
SHLD	Store H and L registers directly in memory
LXI	Load register pair with immediate data
LDAX	Load accumulator from address in register pair
STAX	Store accumulator in address in register pair
XCHG	Exchange H and L with D and E
XTHL	Exchange top of stack with HL

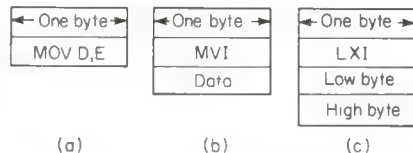
The addition of an X indicates usage of an “extended” 16-bit register, also called a register pair (RP).

In addition to these mnemonics, each transfer instruction must specify a source and a destination register for the data to be moved. Symbolically, the destination register is named first, followed by the source register. For example, a register-to-register move would be shown as MOV (destination),(source). If data in accumulator A is to be moved into the B register, this becomes MOV B,A. This operation does not alter the data in the A register, but merely copies the 8 bits of data into the B register. If the letters are reversed (MOV A,B), the direction of transfer is reversed. Then data is moved from register B to register A.

Data also can be moved to and from memory M with a MOV instruction. If the data byte in register D is to be moved to memory, the instruction is MOV M,D. The destination is memory, and the location in memory is determined by the address programmed into the HL register.

The basic MOV instruction is called a 1-byte instruction because it occupies 1 byte of memory space. Examples of 1-, 2-, and 3-byte instructions are shown in Fig. 18.16. From Fig. 18.16b, the move immediate instruction (MVI) must be followed by the byte of data to be placed in the specified register. For example, MVI A in the first byte and FF in the second byte would put all 1s in the 8 accumulator bits (FF = 11111111).

Figure 18.16c shows a 3-byte instruction LXI, or load an extended (16-bit) register with



**Fig. 18.16** Examples of (a) 1-, (b) 2-, and (c) 3-byte instructions.

Data transfer group			Arithmetic and logic group			Branch control group		I/O and machine control	
<div> <div> <div>MOV</div> <div> <div> <div>A.A 7F</div> <div>E.A 5F</div> </div> <div> <div>E.B 78</div> <div>E.C 59</div> </div> <div> <div>A.D 7A</div> <div>E.E 5B</div> </div> <div> <div>A.H 7C</div> <div>E.L 5D</div> </div> <div> <div>A.L 7D</div> <div>E.M 5E</div> </div> </div> </div> <div> <div>MOV</div> <div> <div> <div>B.A 47</div> <div>E.B 40</div> </div> <div> <div>B.C 41</div> <div>E.D 42</div> </div> <div> <div>B.E 43</div> <div>E.H 44</div> </div> <div> <div>B.L 45</div> <div>E.M 46</div> </div> </div> </div> <div> <div>MOV</div> <div> <div> <div>C.A 4F</div> <div>E.B 48</div> </div> <div> <div>C.C 49</div> <div>E.D 4A</div> </div> <div> <div>C.E 4B</div> <div>E.H 4C</div> </div> <div> <div>C.L 4D</div> <div>E.M 4E</div> </div> </div> </div> </div>	<div> <div>Move (cont)</div> <div> <div> <div>A, byte 3E</div> <div>B, byte 06</div> </div> <div> <div>C, byte 0E</div> <div>D, byte 16</div> </div> <div> <div>E, byte 1E</div> <div>H, byte 26</div> </div> <div> <div>L, byte 2E</div> <div>M, byte 36</div> </div> </div> <div> <div>Immediate</div> <div> <div> <div>A, byte 01</div> <div>B, dble 11</div> </div> <div> <div>H, dble 21</div> <div>SP, dble 31</div> </div> </div> <div> <div>Load/Store</div> <div> <div>LDAX B 0A</div> <div>LDAX D 1A</div> </div> <div> <div>LHLD adr 2A</div> <div>LDA adr 3A</div> </div> <div> <div>STAX B 02</div> <div>STAX D 12</div> </div> <div> <div>SHLD adr 22</div> <div>STA adr 32</div> </div> </div> </div></div>	<div> <div>ADD</div> <div> <div>A 87</div> <div>B 80</div> </div> <div> <div>C 81</div> <div>D 82</div> </div> <div> <div>E 83</div> <div>H 84</div> </div> <div> <div>L 85</div> <div>M 86</div> </div> </div> <div> <div>INX</div> <div> <div>A 8F</div> <div>B 88</div> </div> <div> <div>C 89</div> <div>D 8A</div> </div> <div> <div>E 8B</div> <div>H 8C</div> </div> <div> <div>L 8D</div> <div>M 8E</div> </div> </div> <div> <div>DCR</div> <div> <div>A 97</div> <div>B 90</div> </div> <div> <div>C 91</div> <div>D 92</div> </div> <div> <div>E 93</div> <div>H 94</div> </div> <div> <div>L 95</div> <div>M 96</div> </div> </div>	<div> <div>Increment†</div> <div> <div>A 3C</div> <div>B 04</div> </div> <div> <div>C 0C</div> <div>D 14</div> </div> <div> <div>E 1C</div> <div>H 24</div> </div> <div> <div>L 2C</div> <div>M 34</div> </div> </div> <div> <div>Decrement†</div> <div> <div>A 3D</div> <div>B 05</div> </div> <div> <div>C 0D</div> <div>D 15</div> </div> <div> <div>E 1D</div> <div>H 25</div> </div> <div> <div>L 2D</div> <div>M 35</div> </div> </div> <div> <div>ORA</div> <div> <div>A B7</div> <div>B B0</div> </div> <div> <div>C B1</div> <div>D B2</div> </div> <div> <div>E B3</div> <div>H B4</div> </div> <div> <div>L B5</div> <div>M B6</div> </div> </div>	<div> <div>Logical*</div> <div> <div>A A7</div> <div>B A0</div> </div> <div> <div>C A1</div> <div>D A2</div> </div> <div> <div>E A3</div> <div>H A4</div> </div> <div> <div>L A5</div> <div>M A6</div> </div> </div> <div> <div>XRA</div> <div> <div>A AF</div> <div>B A8</div> </div> <div> <div>C A9</div> <div>D AA</div> </div> <div> <div>E AB</div> <div>H AC</div> </div> <div> <div>L AD</div> <div>M AE</div> </div> </div> <div> <div>ORA</div> <div> <div>A B7</div> <div>B B0</div> </div> <div> <div>C B1</div> <div>D B2</div> </div> <div> <div>E B3</div> <div>H B4</div> </div> <div> <div>L B5</div> <div>M B6</div> </div> </div>	<div> <div>Jump</div> <div> <div>JMP adr C3</div> <div>JNZ adr C2</div> </div> <div> <div>JZ adr CA</div> <div>JNC adr D2</div> </div> <div> <div>JC adr DA</div> <div>JPO adr EA</div> </div> <div> <div>JPE adr FA</div> <div>JP adr F2</div> </div> </div> <div> <div>Call</div> <div> <div>CALL adr CD</div> <div>CNZ adr C4</div> </div> <div> <div>CZ adr OC</div> <div>CNC adr D4</div> </div> <div> <div>CC adr DC</div> <div>CPO adr E4</div> </div> <div> <div>CPE adr EC</div> <div>CP adr F4</div> </div> <div> <div>CM adr FC</div> <div>RET C9</div> </div> </div> <div> <div>Return</div> <div> <div>RET C9</div> <div>RNZ C0</div> </div> <div> <div>RZ C3</div> <div>RIM 20</div> </div> </div>	<div> <div>Stack Ops</div> <div> <div> <div>B</div> <div>D</div> </div> <div> <div>D5</div> <div>H</div> </div> <div> <div>E5</div> <div>PSW</div> </div> </div> <div> <div>POP</div> <div> <div>B</div> <div>D</div> </div> <div> <div>D1</div> <div>H</div> </div> <div> <div>E1</div> <div>PSW*</div> </div> </div> </div>	<div> <div>Input/Output</div> <div> <div>OUT byte D3</div> <div>IN byte D8</div> </div> <div> <div>Control</div> <div> <div>D1 F3</div> <div>E1 F8</div> </div> <div> <div>NOP 00</div> <div>HLT 76</div> </div> </div> <div> <div>New Instructions (9085 Only)</div> <div> <div>RET C9</div> <div>RNZ C0</div> </div> <div> <div>RZ C3</div> <div>RIM 20</div> </div> </div></div>		





## 18-16 Microprocessors and Microcomputers

immediate 2 bytes of data. An example is LXIH, which loads the register pair HL with data (or an address) specified in the next 2 bytes. The low byte for register L must appear in the byte immediately after the LXI instruction, followed by the high byte which is loaded into register H.

Space does not permit complete coverage of the entire instruction set. The manufacturer's programming manual for each microprocessor must be consulted. A few examples are covered for each group of instructions. Table 18.5 summarizes the entire instruction set and op codes needed to write and understand programs for 8085 chip.

**example 18.5** Given the following information in system memory, what bits reside in the D register? Consider address 8000 as the first step of a program.

Address	Data (hexadecimal)	Status
8000H	3E	Op Code
8001	1F	Data
8002	57	Op code

**solution** Here 3E appears at address 8000H. The first program step is always an op code fetch. From Table 18.5, 3E is an op code for MVIA (move immediate data to accumulator). Figure 18.16b specifies an MVI instruction as 2 bytes; therefore, at 8001 (second byte), the digital number 1F is moved into the accumulator. At 8002, op code 57 from Table 18.5 is MOV D,A. This instruction copies the number 1F from the accumulator into the D register. Therefore,

$$1F \text{ (hexadecimal)} = \underbrace{00011111}_{\text{bits in D register}}$$

**example 18.6** Given the 3-byte instruction LXID, what hexadecimal number must be placed in the 3 memory bytes to store 55 in the D register and 44 in the E register?

**solution** LXID from the description under data transfer group is "load register pair with immediate data." This loads the extended register pair D,E with data specified in the next 2 bytes.

### Memory

First byte	11	Op code for LXID from Table 18.5
Second byte	44	Low-order byte in E
Third byte	55	High-order byte in D

**ARITHMETIC AND LOGIC GROUP** This group can be subdivided into two subgroups, A and B.

**A. Arithmetic** Add, subtract, increment, or decrement memory or register data.

- ADD Add to accumulator
- ADI Add immediate data to accumulator
- ADC Add to accumulator, using carry flag
- ACI Add immediate data to accumulator, using carry flag
- SUB Subtract from accumulator
- SUI Subtract immediate data from accumulator
- SBB Subtract from accumulator, using borrow (carry) flag
- SBI Subtract immediate from accumulator, using borrow
- INR Increment specified byte by 1
- DCR Decrement specified byte by 1
- INX Increment register pair by 1
- DCX Decrement register pair by 1
- DAD Double register add: Add contents of register pair to H,L register pair

**example 18.7** The number 1A is stored in the accumulator. Using an arithmetic instruction, make the accumulator contents equal 00H.

**solution** One solution would use SUB A, which subtracts the contents of register A from register A. The result is placed in the accumulator:  $1A - 1A = 00H$ . If the difference were not zero, the accumulator bits would be presented in a 2's-complement format.

Another solution is SUI, which subtracts the next immediate byte from the accumulator. Instruction SUI works only with the accumulator.

Byte 1	D6	op code for SUI
Byte 2	1A	Data to be subtracted

This solution uses 2 bytes, whereas the first solution requires only 1 byte. Both solutions accomplish the objective; however, the first saves time and memory space.

**B. Logic** Perform logical, compare, and rotate operations on the data.

ANA	Logic AND with accumulator
ANI	Logic AND with accumulator, using immediate data
ORA	Logic OR with accumulator
ORI	Logic OR with accumulator, using immediate data
XRA	Exclusive logic OR with accumulator
XRI	Exclusive OR, using immediate data
CMP	Compare 8-bit value with accumulator
CPI	Compare, using immediate data
RLC	Rotate accumulator left
RRC	Rotate accumulator right
RAL	Rotate left through carry
RAR	Rotate right through carry
CMA	Complement accumulator
CMC	Complement carry flag
STC	Set carry flag

**example 18.8** The ANA instruction performs a logic AND operation with a specified register. Given the ANAB instruction with the following bits in the accumulator and register B, list the resulting bits in the accumulator.

	D7	D0	
Accumulator =	1 0 1 0 1 0 1 0		(AAH)
Register B =	1 1 0 0 1 1 0 0		(CCH)

**solution** Refer to Chap. 14. An AND gate operation yields a 1 output if both inputs are 1 and a 0 output if the inputs are different. All the bits in registers A and B are ANDed one at a time:

A =	10101010	(AAH)
B =	11001100	(CCH)
Result in A =	10001000	(88H)

**example 18.9** The CMP instruction compares the accumulator bits with a specified register and sets flags to indicate the result, as shown in Table 18.6. Given a CMPC instruction with CFH in the accumulator and CDH in register C, will the zero (Z) and carry (CY) flags be set?

**TABLE 18.6 Results of Compare Operation**

	Flag	Condition	Meaning
If sign flag is 1 (positive number)	Zero	Set (1)	Equal values in A and register
	Zero	Not set (0)	Unequal values in A and register
	Carry	Not set (0)	A $\geq$ register value
	Carry	Set (1)	A < register value
If sign flag is 0 (negative number) or one value is complemented	Zero	Set (1)	Equal values in A and register
	Zero	Not set (0)	Unequal values in A and register
	Carry	Not set (0)	A < register value
	Carry	Set (1)	A $\geq$ register value

## 18-18 Microprocessors and Microcomputers

**solution** Since  $A > \text{register C}$  ( $CF > CD$ ), from Table 18.6  $Z = 0$  (not set) and  $CY = 0$  (not set).

**example 18.10** Assume the accumulator contains the bit pattern 10001000. If you use an RLC instruction, how many times must this instruction be executed to produce 22H in the accumulator? Note: RLC rotates A contents 1 bit to the left with the MSB transferring to the LSB.

**solution** The transfer sequence after each execution is

Initial value:	10001000	(88H)
First execution:	00010001	(11H)
Second execution:	00100010	(22H)

Therefore, two executions are required.

Examination of the arithmetic and logic group instructions shows the accumulator to be central in the execution of these instructions. Remember the ALU in the CPU always uses data supplied from the accumulator. Compare instructions compare an 8-bit value with the accumulator. Rotate instructions shift the accumulator contents 1 bit position to the left or right. Instructions included in this group, not limited to the accumulator, are the decrement and increment instructions which can be applied to other registers as well. Most of these instructions affect different combinations of the five flag bits often used as indicators of the operational results. The programming manual should be consulted for flag status information stemming from many possible combinations.

**BRANCH CONTROL GROUP** Changes in the normal program sequence flow are initiated with branch control instructions. Two categories, unconditional and conditional, make up this group.

**Unconditional** Decision to jump is not based on flag conditions.

JMP Jump to specified address  
CALL Call subroutine at given address  
RET Return to previous program

**Conditional** Decision to jump is based on flag conditions. Terminology is as follows with the effect on condition flags shown in parentheses:

NZ Not zero ( $Z = 0$ )  
Z Zero ( $Z = 1$ )  
NC No carry ( $C = 0$ )  
C Carry ( $C = 1$ )  
PO Parity odd ( $P = 1$ )  
PE Parity even ( $P = 1$ )  
P Plus ( $S = 0$ )  
M Minus ( $S = 1$ )

Using this terminology, we have

Jump	Call	Return
JC	CC	RC (Carry)
JNC	CNC	RNC (No carry)
JZ	JNC	RZ (Zero)
JNZ	CNZ	RNZ (Not zero)
JP	CP	RP (Plus)
JM	CM	RM (Minus)
JPE	CPE	RPE (Parity even)
JPO	CPO	RPO (Parity odd)

Two additional branching instructions change the program counter:

PCHL Moves H and L to PC.

RST Special restart instruction used with interrupts. Restart is a special-purpose call.

Refer to Sec. 18.6 for an example of an unconditional jump. Examples of CALL and RET instructions are given under the section "Subroutines" (see Fig. 18.20).

**STACK, I/O, AND MACHINE CONTROL GROUP** The fourth group of instructions includes those for stack and I/O operations, and for machine controls.

**Stack** Affects stack and stack pointer.

**PUSH** Push 2 bytes of data onto the stack

**POP** Pop 2 bytes of data off the stack

**XTHL** Exchange top of stack with H and L

**SPHL** Move contents of H and L to stack pointer

#### I/O

**IN** Initiate input operation (from data bus to the accumulator)

**OUT** Initiate output operation (from accumulator to data bus)

#### Machine controls

**EI** Enable interrupt system

**DI** Disable interrupt system

**HLT** Halt

**NOP** No operation

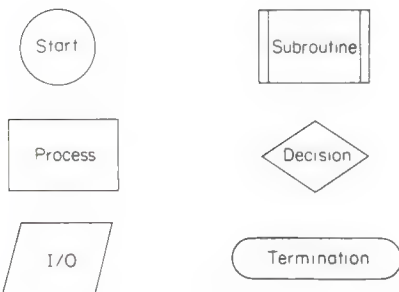
## 18.6 SOFTWARE—WRITING A PROGRAM

Writing programs for microprocessor systems is an art learned by doing rather than reading. A few tools and general techniques are discussed in this section; however, the same end result may be accomplished by using different instructions. If memory space is at a premium or time is essential, the shortest possible program must be developed.

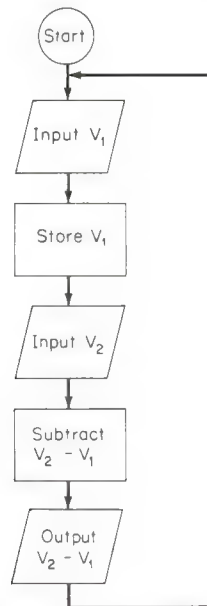
### STEPS IN WRITING A PROGRAM

**Step 1: Description** Write down in plain language what the program is to accomplish. A detailed description at the beginning will greatly facilitate the mechanics involved in writing the program. This also assists others in understanding the program, in addition to serving as a future reference to the programmer.

**Step 2: Flowchart** A flowchart is a pictorial diagram of the overall program flow. Figure 18.17 shows symbols often used to draw charts. Consider a program designed to input two numbers from a digital voltmeter (DVM), subtract them, output the result to a controller, and repeat the process (Fig. 18.18). The circle indicates the starting point. The parallelogram block inputs the first voltage from the DVM to be stored (a process depicted by the



**Fig. 18.17** Flowchart symbols.



**Fig. 18.18** Example of a flowchart.

rectangle) in the microprocessor. Next, V2 is inputted to the accumulator, V1 is subtracted from V2 in the ALU unit, and the difference is sent to an output port for use by a controller. By jumping back to start, the program runs in a continuous loop.

**Step 3: Examine system addresses** First, the memory map of the processor system must be examined to determine the addresses available for programming and data storage.

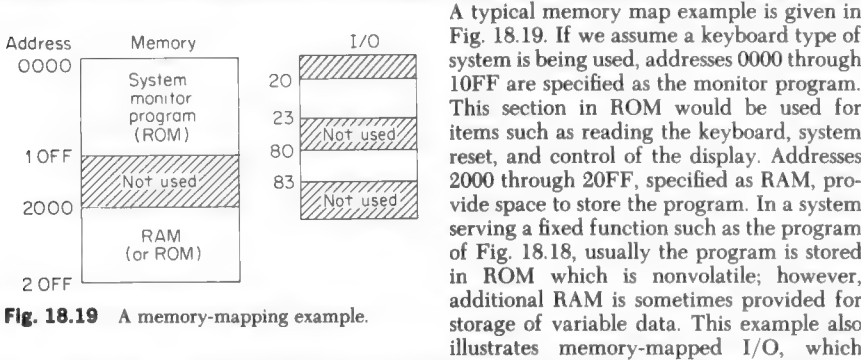


Fig. 18.19 A memory-mapping example.

means the I/O ports are identified by port addresses. IN and OUT instructions are followed by the port address.

**Step 4: Instruction set** With the flowchart as a guide, instructions from the instruction set can be selected to accomplish the program objectives. An assembly language program can be written now by using the mnemonics for the microprocessor. The mnemonics are written opposite each address along with a remark concerning the purpose of the instruction.

**Step 5: Coding** Once the mnemonics are selected, the corresponding op codes from the manufacturer's programming sheet are written by the mnemonic.

Table 18.7 illustrates how steps 1 to 5 are put into practice for the program of Fig. 18.18. Each IN (or OUT) instruction is followed by the port address at program addresses 2001, 2004, and 2007. At address 2001, V1 is in the accumulator. Since the next IN instruction (2003) will input V2 into the accumulator, V1 must be stored in another location. This is accomplished by MOV B,A, which moves V1 from the accumulator to register B (any other register could be used). After V2 is in the accumulator, the SUBB instruction subtracts the B register contents (V1) from the new accumulator value (V2), writes the result back into

TABLE 18.7 Program for Fig. 18.18

Step 1			
Program to input V1 from port 20 and V2 from port 21; subtract two voltages; output difference to port 80; loop.			
Step 2			
Step 2 is the flowchart in Fig. 18.18.			
Step 3, Address	Step 4, Mnemonic	Step 5, Op code	Remarks
2000H	IN	DB	Input V1 from port
2001	20	20	Port 20
2002	MOV B,A	78	Store V1 in B register
2003	IN	DB	Input V2 from port
2004	21	21	Port 21
2005	SUB B	90	V2 - V1
2006	OUT	D3	Output difference to
2007	80	80	Port 80
2008	JMP	C3	Jump back to start
2009	00	00	Adr low byte
200A	20	20	Adr high byte

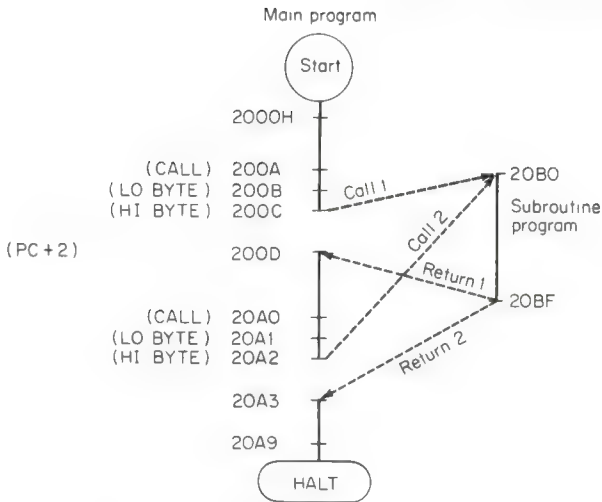


Fig. 18.20 Execution of a subroutine.

the accumulator, and outputs the difference to port 80. JMP at address 2008 unconditionally jumps the program back to the starting address 2000 (low byte 00, high byte 20), causing the program to run in a continuous loop.

**SUBROUTINES** Subroutines are used in programs requiring a certain operation to be performed several times during the main program. A subroutine is assigned addresses outside the main line program and called when needed.

Figure 18.20 shows a main program starting at address 2000H. At 200A a CALL instruction appears followed by the subroutine address B0 (low byte first) and 20 (high byte second). This 3-byte instruction saves the next program counter address following the CALL plus 2 on the stack (200D) as a return address. Programming RET at the end of the subroutine forces (pops) the saved address off the stack, thus jumping the PC back to the main program. A return instruction does not require a specified address because the stack takes care of this task automatically. The same procedure is used at 20A0 with the same subroutine. Advantages of using subroutines are that they save memory space, simplify writing programs, and encourage a modular program structure.

It is also possible to CALL a subroutine within another subroutine, as illustrated in Fig. 18.21. When subroutines are called within subroutines, it is called *nesting*. Return addresses are saved on the stack in the correct order to keep the correct program sequence. The stack is sometimes referred to as a LIFO (last in, first out); hence, an address at point A on the diagram is placed on the stack first, and B is pushed onto the stack last. After completion of subroutine 2, address B is popped off the stack first, allowing subroutine 1 to resume. At the end of subroutine 1, address A is popped off the stack, and there is return to the main program.

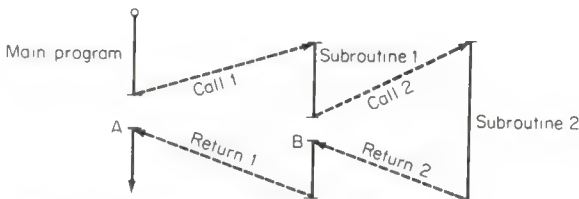


Fig. 18.21 Nesting of subroutines.



## 18.7 SUPPORT DEVICES

Development of a microcomputer requires selection of support devices for the microprocessor. The number of devices used in a system will depend on the functions that the system must provide. A few devices are mentioned in this section.

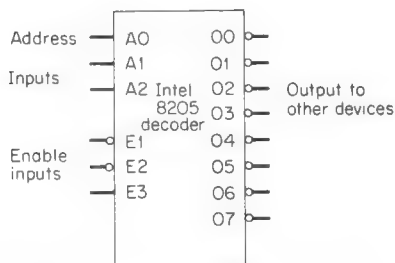
**DECODERS** When several memory or I/O chips are used in a system, a means of activating each device must be provided. Decoders act as a switch to enable one of several devices in a system. Figure 18.22 shows the logic symbol for a "1 out of 8" binary decoder. Signals on A0-A2 (address inputs) and E1-E3 (enable inputs) will produce an output on only one of the decoded output lines (O0-O7). The outputs are used as chip enable (CE) or chip select (CS) signals to activate other devices. Manufacturers publish truth tables showing input versus output data.

**I/O PORTS** Refer to Fig. 18.23. This device is enabled by selecting a digital control word consisting of STB (strobe), DS1 and DS2 (device select), and MD (mode). Terminals STB, DS1, DS2, and MD comprise the control logic used to set up device selection, data latching, output buffers, and/or an interrupt request mode. Normally an I/O port is initialized in the beginning of a program to configure the circuitry as an input or output. Then the initialized device will latch into this configuration and will not have to be reprogrammed unless the same port is used for the opposite function later in the routine. Manufacturers supply truth tables in their specifications for setup configurations.

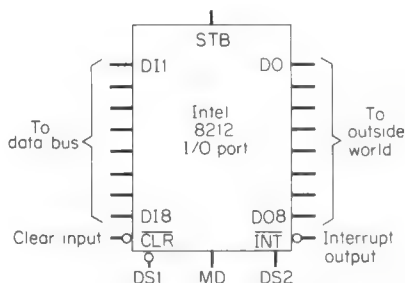
**PROGRAMMABLE PERIPHERAL INTERFACE (PPI)** Support devices like the PPI offer a system great flexibility in a small package. Software controls I/O ports in various combinations. A block diagram of the Intel 8255A PPI is shown in Fig. 18.24. Processors are interfaced to peripheral devices via the three-state bidirectional data bus buffer. Transmission and reception of data are controlled by the read-write control logic block programmed by the CPU. As a result, the three 8-bit ports (A, B, C) can be selected as inputs or outputs for the microprocessor. Nomenclature for Fig. 18.24 is shown in Table 18.8.

**OTHER SUPPORT DEVICES** Many other support chips are used in different systems. The following list summarizes the characteristics of some chips.

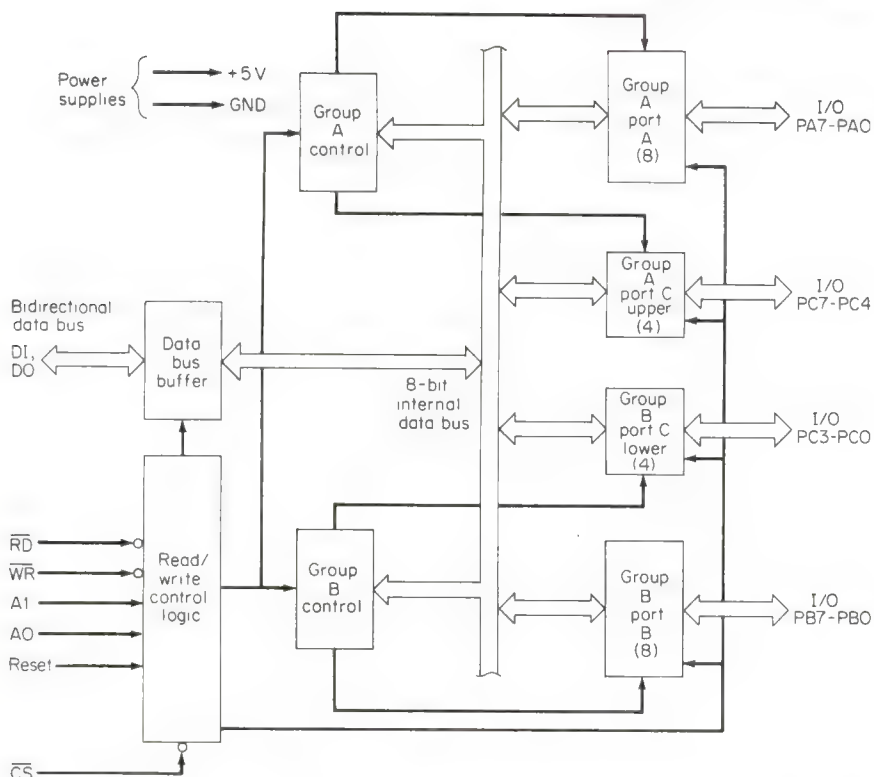
1. Programmable DMA (direct memory access) controller. Certain peripheral devices operate at higher speeds than the CPU or may require use of the bus system independent of the CPU. A DMA controller provides signals to tristate the CPU from the bus via a HOLD function, generate timing and control signals, direct data transfer between peripherals, and return control to the CPU at completion of DMA activity.
2. Programmable interrupt controller. This device permits several vectored interrupts to be programmed if the microprocessor does not have the necessary capability.
3. Programmable CRT controller. Systems requiring a CRT display use this device to interface and control CRT circuitry via the CPU.
4. Programmable keyboard interface. A keyboard interface will scan a keyboard used to program a processor system and send the result to the CPU.



**Fig. 18.22** A "1 out of 8" binary decoder. (Courtesy Intel Corporation.)



**Fig. 18.23** Programmable 8-bit I/O port. (Courtesy Intel Corporation.)



**Fig. 18.24** Block diagram of an 8255A programmable peripheral interface. (Courtesy Intel Corporation.)

The manufacturer's data must be consulted to understand the individual support devices, to troubleshoot systems, or to design interfaces.

## 18.8 MICROPROCESSOR SURVEY

Our discussion now shifts from the Intel 8085 to a brief description of other types. Space does not permit exhaustive coverage of every processor; however, an introduction to the programming structure, including registers and instructions, is presented. Before delving into a survey, we discuss addressing modes used by various CPUs.

Three addressing modes were covered earlier (immediate, direct, indirect). Three additional categories of indirect addressing—extended, implied, and accumulator—are described now along with indexed addressing and its special case, called *relative addressing*.

**EXTENDED ADDRESSING** Extended addressing is a type of indirect addressing that uses a 16-bit register and a 3-byte instruction. Byte 1 contains the instruction, byte 2 contains 1

**TABLE 18.8** Nomenclature for Fig. 18.24

D7-D0	Data bus	A0, A1	Port address
Reset	Reset input	PA7-PA0	Port A bits
$\overline{CS}$	Chip select	PB7-PB0	Port B bits
$\overline{RD}$	Read input	PC7-PC0	Port C bits
$\overline{WR}$	Write input	Vcc	+5 V

byte of the address, and byte 3 contains the second byte of the address. Low and high address bytes must be placed in the proper sequence as determined by the type of microprocessor being used.

**IMPLIED ADDRESSING** Implied addressing is another type of indirect addressing mode in which the instruction op code contains the operand address. An example would be the 8085 MOV instructions or STC affecting only the carry flag.

**ACCUMULATOR ADDRESSING** Accumulator addressing is a special case of implied addressing. The instruction relates directly to an accumulator which contains the operand. Examples from the 8085 instructions include decimal adjust accumulator (DAA) and rotate instructions.

**INDEXED ADDRESSING** Indexed addressing employs an "offset number" plus a specified address which point to another address storing the operand. The process works as follows:

1. The index register contains an address. Assume 3000H as an example.
2. An offset number is specified in the second byte of the instruction, for example, 2AH.
3. The program counter jumps to the sum  $3000H + 2AH$ , or 302AH.

The offset, sometimes called *displacement*, is an 8-bit value limited in addresses from 00H to FFH. If a number of operands are stored sequentially in memory, indexed addressing provides an easy method of advancing the program by incrementing the index register.

**RELATIVE ADDRESSING** Relative addressing is a special case of indexed addressing in which the program counter is used as the index register. The program counter points to the displacement stored in memory, adds that offset to its own value, and points to the resulting address storing the operand.

**COMPARISON OF MICROPROCESSORS** The 8080, Z80, 6800, and 6502 provide a cross section of popular 8-bit processors using different programming and hardware configurations.

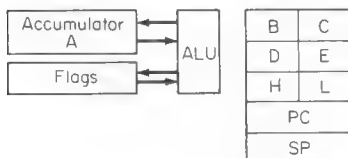


Fig. 18.25 Simplified block diagram of the 8080 microprocessor.

A simple block diagram showing registers available to a programmer along with a brief description of operation is given for each type.

**Intel 8080 (Fig. 18.25)** In previous sections of this chapter we described the 8085 in detail; the 8080 was the forerunner of the 8085. The basic differences between the two devices are shown in Table 18.9. Programs written for the 8080 use the same instructions as the 8085 except for the additional 8085 features of interrupts and serial data I/O.

**Z80** See Fig. 18.26. Examination of the register configuration shows some similarity to the 8080/8085; however, additional registers include duplicate accumulators, flags, extended register pairs B'-D'-H', and two index registers. The added capability permits indexed addressing as well as numerous data moves between registers. Data words also can be rotated

TABLE 18.9 Intel 8080 Compared with Intel 8085

8080	8085
One built-in interrupt	Five priority interrupts available
No SOD or SID	Serial I/O ports included
Needs external two-phase clock chip	Built-in clock generator
+12-, +5-, -12-V power supplies	Single +5-V power supply
2- $\mu$ s machine cycle	Approximately 50% faster
78 instructions	Added instructions for interrupts and serial I/O
Needs external controller chip	Built-in system controller

in any register or memory location. Extensive memory-block move and search commands provide flexible data transfers. Seven flag bits are available for testing program status. The instruction set of Z80 has about 158 instructions, which include the 8080 instructions under different mnemonics.

**6800** An absence of scratchpad registers is noticeable in Fig. 18.27. Instead of storing and moving data inside the processor, the 6800 stores and moves data in RAM memory with various instructions. This CPU is an example of a *memory-oriented* type of processor. Short instructions place the first 256 memory locations within easy access of the processor. Seven addressing modes previously described are available. Many instructions require a selection of the addressing mode as well as the operation to be performed. Index and relative addressing provides efficient movement of data between memory locations.

All data in the internal registers and accumulators are 8 bits wide. A condition code register of 6 bits serves the same function as the 8085 flags. The bus structure of the 6800 contains a 16-bit address bus and 8-bit data and control buses. Three software interrupts and one hardware interrupt also are supplied; however, this is not a vectored interrupt as in the 8085. A two-phase external clock chip must be used with the 6800. The newer type 6802 has an on-chip clock generator and an on-chip RAM of 128 locations 8 bits wide.

**6502** See Fig. 18.28. As in the case of the 6800 family, the 6502 is memory-oriented. Only one 16-bit register, the program counter, is present. All other registers are 8 bits wide, including an accumulator, status register (flags), stack pointer, and two index registers.

An outstanding feature of the 6502 is the 13 possible addressing modes. The modes that can be specified for the 56 instructions of the instruction set include the addressing modes discussed earlier—immediate, accumulator, implied, relative, and indirect. In addition to these, the following addressing modes are also available:

1. Absolute. Similar to extended addressing on the 6800. Bytes 2 and 3 contain the 16-bit memory address where the operand is located.

2. Zero page. Similar to direct addressing in the 6800, sometimes called *base-page* addressing. The CPU categorizes 256 bytes of memory into a page. Zero page refers to memory addresses 00H to FFH, the first 256 memory bytes. So 1K byte of memory would have four pages.

3. Indexed indirect X. A zero page address in the second byte is the offset value added to index register X resulting in another zero page address. This address holds yet another address pointing to the operand anywhere in memory. Another term used for this mode is *preindexing*.

4. Indirect indexed Y. The second instruction byte specifies a zero page address. Immediately following the zero page address byte is another address byte (16-bit indirect address) used to fetch yet another 8-bit address which is added to the Y index register, resulting in a pointer for the operand. This is also referred to as *postindexing*.

5. Zero page X and zero page Y. Both modes are types of indexed addressing. The second instruction byte address is added to the index register (X or Y) to get the effective address. This sum must be less than the zero page limit of FFH.

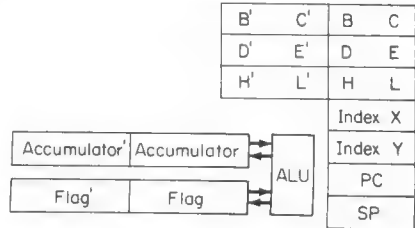


Fig. 18.26 Simplified block diagram of the Z80 microprocessor.

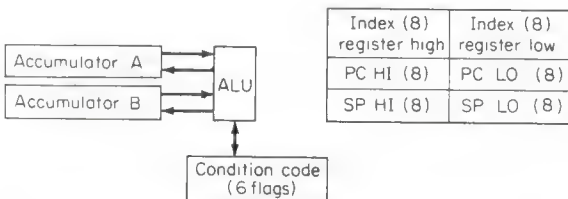
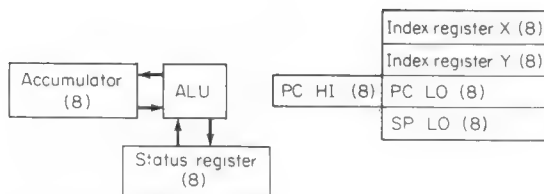


Fig. 18.27 Simplified block diagram of the 6800 microprocessor.



**Fig. 18.28** Simplified block diagram of the 6502 microprocessor.

6. Absolute X and absolute Y. Another type of indexed addressing is used. The index register (X or Y) is added to the address in the second and third instruction bytes, providing a fetch address for the operand. The sum must be on zero page.

## 18.9 SUMMARY

This chapter provides a basic introduction to microprocessor operation. Technological advancement continues at an astonishing pace with the development of 16-bit microprocessors, multi-microprocessor systems, and single-chip microcomputers. The study of any microprocessor requires the user to do the following:

1. Examine the block diagram (architecture).
2. Study the instruction set.
3. Relate the instruction set to the architecture.
4. Review the timing diagrams and control signals.

Development systems supplied by manufacturers provide a tremendous learning tool for hardware and software. These systems come in many sizes, shapes, and prices. Another valuable learning tool is a logic analyzer (Chap. 19) which presents a visual readout of programs, control signals, and timing relationships.



# Chapter 19

## Logic Analyzers, Logic Probes, and Signature Analysis

### “Logic Analyzers” by Pat Zagaria

*Product Marketing Manager*

*Philips Test & Measuring Instruments, Inc.*

*Mahwah, New Jersey*

### “Signature Analysis” by Kenneth Jessen

*Material Engineer*

*Hewlett-Packard Co.*

*Loveland, Colorado*

#### 19.1 INTRODUCTION TO LOGIC ANALYZERS

More and more in our electronic world, digital solutions are used to realize sophisticated functions. The exploding application field of microcomputers is just one example of this fast-growing innovation. But analog test and measuring instruments are not always adapted to the measuring requirements of digital environments. So a new type of instrument, a specialized tool for this digital environment called the *logic analyzer*, has been developed.

In this section we explain why we need a logic analyzer and how it works. We cover the following topics in turn:

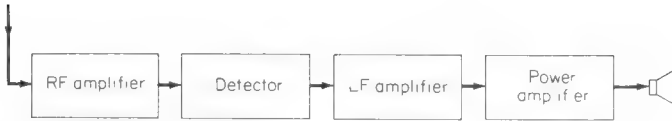
- Brief review of digital circuitry
- Fault conditions in digital circuits
- Basic principles of logic analyzers
- Main functions of logic analyzers
- Applications of logic analyzers
- Glossary of logic analyzer terms

#### 19.2 REVIEW OF DIGITAL CIRCUITRY

First, let us look at an analog circuit. Figure 19.1 shows a very simple, straightforward radio receiver. The signal reaching the input is processed sequentially (we could also say serially in time) by all the stages in turn, finally arriving at the output. Transfer of information tends to be over one line. Fault finding in such a system is simple. It can be executed by just studying every stage separately with a simple measuring tool or an oscilloscope. For example, if the signal at the input of the detector stage is correct but that at the output of this stage is incorrect, then the detector stage must be defective. This *sequential* processing of the signal is the most significant property of most analog circuits.

Now let us look at a simple digital circuit. Figure 19.2a shows an AND gate with three

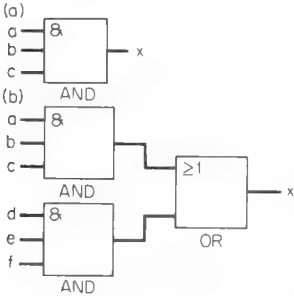




**Fig. 19.1** Sequential processing of analog signals.

inputs and one output. The three input conditions *a*, *b*, and *c* of this gate together determine the output condition; so for fault finding on this gate, we need an instrument with four inputs, which will measure and/or display all four signals at the same time.

The second AND gate in Fig. 19.2*b* adds three more input conditions to the circuit. Combining these two AND gates via an OR gate, we get a basic circuit often found in digital applications.

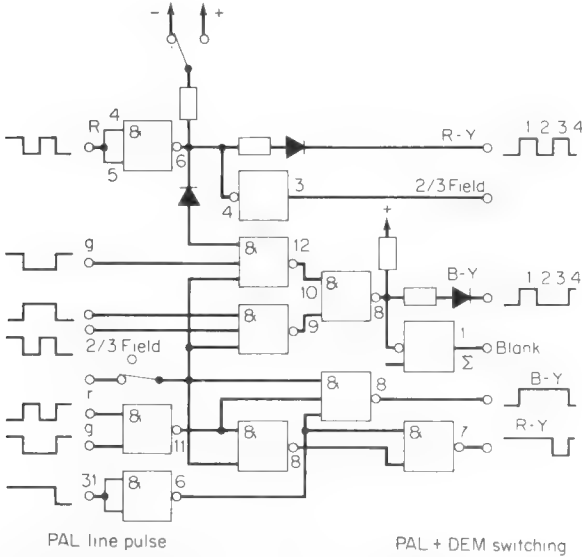


**Fig. 19.2** Combinatorial logic circuit.

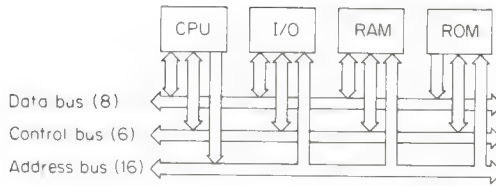
If we want to trace the signals in such a circuit to find a fault, we need to be able to trace at least seven signal lines; and if we want to see the intermediate signals as well, we need two extra inputs. Only the capability to trace these nine signal lines at the same time enables us to find the fault. This is what we call *combinatorial* circuitry. Digital circuits have multinodal signal points. This is a major difference between analog and digital circuits. In practical applications however, not only are the digital circuits combinatorial; also the signal is sequentially processed.

This combination of sequential and combinatorial signal processing is clearly illustrated in Fig. 19.3. On the left, we see several input conditions. Only if these input conditions are fulfilled during a certain time will the resulting output condition be correct.

A more complicated example of such a system is the microcomputer system of Fig. 19.4. The signals (data) are sequentially transported over the bus lines. Functions such as the central processing unit and the memories are realized by combinatorial logic. Clearly many



**Fig. 19.3** Combination of sequential and combinatorial signal processing in digital system.



**Fig. 19.4** Microcomputer system.

conditions will have to be examined in combination and in sequence to trace a fault in such a system.

We have the data bus with data present on 8 lines, the 6-line control bus, and the 16-line address bus. In addition there are various output, control, and input lines. All together, there is much more than we will ever be able to examine with an oscilloscope. We need a special, new kind of tool for analyzing all these signals. This is the logic analyzer.

### 19.3 TYPES OF FAULTS IN DIGITAL SYSTEMS

The common faults that can occur in digital circuitry may be due to the hardware, the software, or both; software faults occur, of course, only if the application is controlled by a software program. The following hardware faults may be distinguished:

- No data or wrong data
- Glitches
- Spikes
- Races
- Timing error
- Ringing
- Wrong level

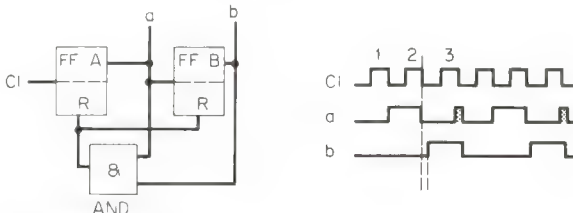
These faults can occur in analog circuits also; however, the large number of signal lines in combinatorial circuits can greatly complicate the task of fault finding.

Software faults are mainly program errors from different sources, or memory defects.

#### HARDWARE FAULTS

**No data** Little need be said about this point. No data or wrong data can be due to a broken line, short circuit, or power failure in some circuits.

**Glitches** Figure 19.5 shows one way in which a glitch can be generated. The circuit is a 4-bit counter gated in such a way that it counts only up to 3. At count 3, both outputs *a* and *b* are high; the AND gate is then enabled, and both flipflops are reset to zero. This means that the *a* output is high only for a very short time here before it is reset. The resulting short pulse (shown in the figure) is called a *glitch*. This type of unwanted signal normally is found only in the design stage. Correct design should eliminate all glitches.



**Fig. 19.5** Glitches.

**Spikes** A spike is an unwanted signal very similar to a glitch. Figure 19.6 shows an example of spike generation. The sharp rising edge of a signal transition present on one or

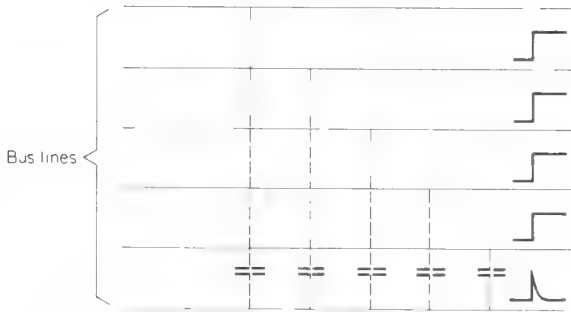


Fig. 19.6 Generation of (parasitic) spikes.

more of the lines in a bus system can cause a parasitic signal to appear on one or more bus lines owing to capacitive coupling. This parasitic signal is called a *spike*. It can create problems in a logic circuit connected to the line involved, for the circuit would receive the unwanted signal for a short time.

**Races** Races are quite similar to glitches and spikes. They generally occur when signals of different speed are combined in one logic circuit (see Fig. 19.7). If both signals *a* and *b*

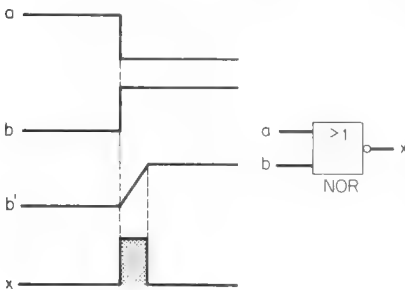


Fig. 19.7 Races.

have sharp edges, as shown in the top half of the figure, there will be no signal at the output of the NOR gate. However, heavy loading of the *b* signal line can make the rising edge of the *b* signal somewhat slower than the falling edge of the *a* signal. Consequently, it will take longer than usual for *b* input to reach its switching threshold. For a short time, therefore, both *a* and *b* will be within the range characterized as low, giving a high output. As soon as the *b* signal reaches the threshold level, the output goes low, which results in a short unwanted output pulse.

**Timing errors** Another hardware fault condition can be caused by timing errors.

Figure 19.8 shows data on a data bus together with the corresponding sampling pulse. Suppose the data are taken from the bus and sent to another device at the falling edge of the sampling pulse. If for some reason the data sampling pulse is delayed and appears at the moment when the data are changing (indicated by the crossing of the data lines), then undefined data are sampled and the result may be false.

**Ringing** Ringing is another hardware fault. Figure 19.9 shows a two-input AND gate. If the inductance of one of the input lines of the AND gate is excessive, the sharp edge of

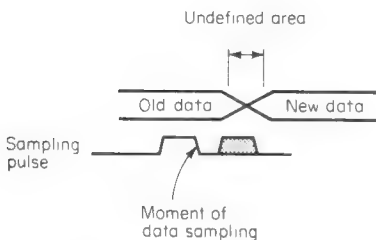


Fig. 19.8 Timing error.

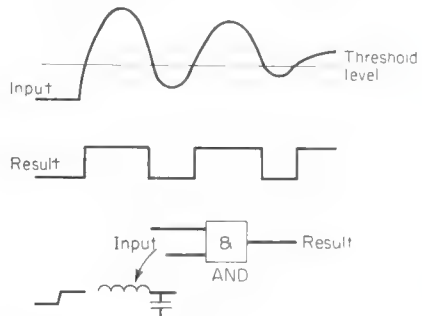


Fig. 19.9 Ringing.

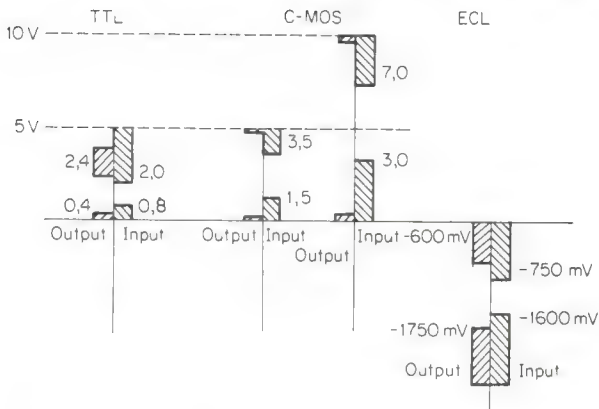


Fig. 19.10 Logic levels.

an input signal could cause oscillations of such an amplitude that the signal passes the gate threshold several times, causing the output shown.

**Wrong level** Each family of logic circuits has its own defined range of switching levels. Figure 19.10 shows the logic ranges of some popular digital circuits: TTL, CMOS, and ECL.

Let us take a TTL circuit as an example. A TTL output will deliver an output level of less than 400 mV when the signal is logic 0 and of at least 2.4 V when the input signal is logic 1. At the input, however, the circuit will still recognize 800 mV as a 0 and signals from 2 V onward as 1. If the input level is between 0.8 and 2.0 V, the logic condition is undefined. The output signal then can be either a 1 or 0, so a faulty condition can occur. These wrong levels can be caused by excessive loading of the logic gates or by interference.

**SOFTWARE FAULTS** Software errors are mainly program errors. These can be subdivided into three main groups: wrong instructions, latent faults, and timing faults.

**Wrong instructions** The first is an error caused by wrong instructions, so the program is not executed as planned. This could also be due to missing instructions, wrong addresses, etc.

**Latent faults** Figure 19.11 shows a small part of a program where an add operation is followed by an unconditional jump instruction. If now the jump instruction is changed by mistake into a "jump nonzero" instruction, the program will run as planned as long as the result of the preceding "add b" operation is not a zero. However, if the result is a zero, the jump instruction will not be executed and the program will continue with the output routine, which certainly was not the intention. Errors of this type may be due to wrong instructions in the program or even to hardware defects. As we can see from Fig. 19.11, the codes in this example for the jump and the jump nonzero instructions differ by only 1 bit, the last.

**Timing faults** In some programs for communication between processor and peripherals, the transmission rate is software-controlled. If the peripheral unit is too slow in executing its function and the software is not programmed to wait for completion of the operation in question, the program will continue, which means that various faults can occur.

**Memory defects** Another type of error, sometimes wrongly regarded as a software fault, is a memory defect that is due, for example, to an open circuit or a short circuit in the memory lines. Such a memory defect also could be the cause of the mix-up between the

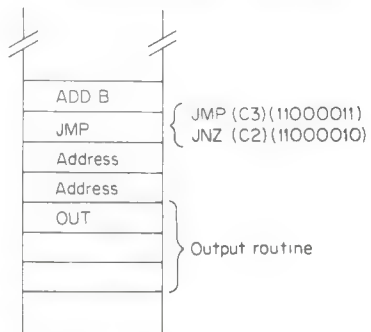


Fig. 19.11 Latent software fault.

unconditional jump and jump nonzero instructions mentioned above. In fact, these are hardware defects, and we need hardware tools to find them.

Our short review of digital circuits and the possible faults makes it quite clear how difficult it can be to trace a fault in such circuits, because of the complexity and speed of the hardware and the intricate structure of the software. This is particularly true if more than one fault is present.

We should now be able to specify the requirements of the ideal tool for these applications, the logic analyzer.

#### 19.4 REQUIREMENTS ON LOGIC ANALYZER

First we list some general requirements:

- We should have the capability to trace and display many input signals, say 16, at the same time. This will permit measurements on combinatorial logic circuitry.
- We require an optimum sampling rate, to be able to work with logic circuits of all kinds.
- The input level (threshold) should be adaptable to different logic families.
- Storage facilities are needed, to capture sequences of data occurring only once or at very low repetition rates.
- We should have extensive trigger facilities, enabling us to capture specific blocks of data in a data stream.
- And we need a variety of convenient display modes to enable us to analyze the problem on hand simply.

A more specific requirement is the capability to measure in the data domain as well as in the time domain. The *time domain* is the field for work in the hardware environment.

- Here we work with a high clock rate, normally asynchronously, for optimum resolution in high-speed circuitry.
- We should have detection facilities for hardware faults such as glitches, spikes, races, and ringing.
- We should also be able to make accurate time measurements.
- In the *data domain*, which is mainly in the software environment, we normally work with a synchronous system clock taken from the system under test.
- We also should be able to compare the incoming data with reference data stored in additional memory.
- Selection of specific data in a data stream should be possible via selective sampling.
- It should be possible to present captured data in a state table in various formats: binary, octal, or hexadecimal (hex). Presentation in the map mode is also useful for solving special problems.

As an additional feature, logic analyzers should be able to offer facilities for analog measurement of quantities such as level and time with an oscilloscope (which is still the only tool for detailed real-time measurements).

After this brief discussion of why we need logic analyzers and what they should be able to do, it is time to consider what this invaluable measuring tool really is and how it does what we need it to do.

#### 19.5 BASIC PRINCIPLES OF THE LOGIC ANALYZER

As may be seen from Fig. 19.12a, a logic analyzer is basically nothing more than a *memory*, which functions here as a multiple shift register for at least 16 bits in parallel, where we store digital *data*, taken from a system under test. The operation of the logic analyzer is controlled by a *clock signal* (the clock together with the data input is the data acquisition unit). A sample of the data present is taken each time a clock pulse occurs and transferred to the memory. How does this work?

After the first clock pulse, one data word is stored in the first memory location; see Fig. 19.12b. On receipt of the next clock pulse, this data word is shifted one place further, and the next data word is transferred to the first memory location. Each subsequent clock pulse causes the string of data words to be shifted one place further in this way. Figure 19.12c is a photograph of a logic analyzer.

**MEMORY** A prime consideration for defining the memory of a logic analyzer is the width of our data word, or how many bits in parallel can be taken at every clock pulse.

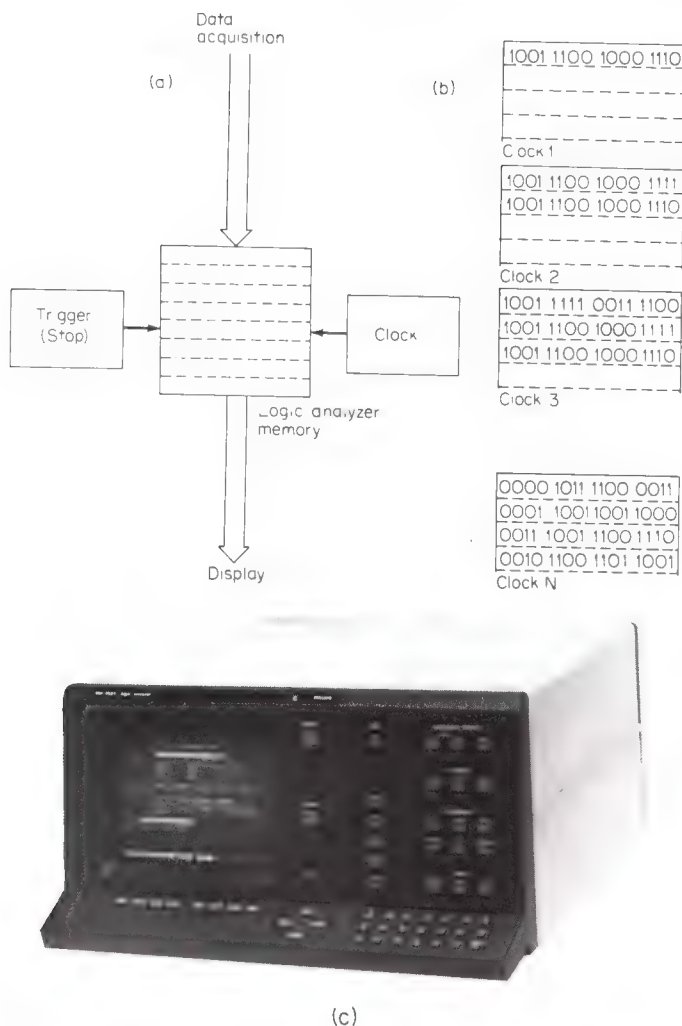


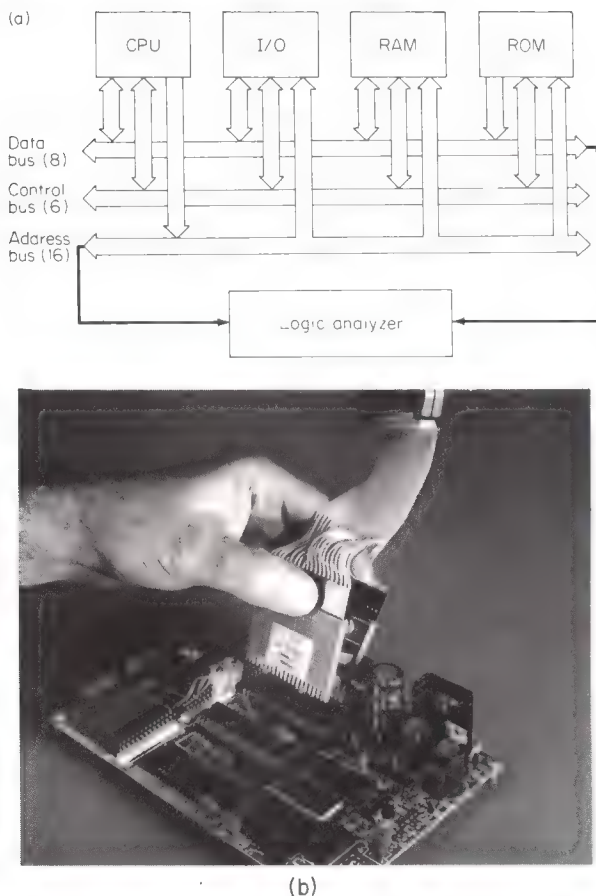
Fig. 19.12 Basic design of logic analyzer.

In accordance with the requirements mentioned previously, Philips logic analyzers have 16 or more inputs (up to 96 channels) so the data word we can take from the system under test is at least 16 bits wide. The number of words we can store at a time is a second consideration. It may be 64; then the memory will have 64 locations for the data words—but any other size is possible in principle. Sizes up to 2000 are not uncommon.

We have seen how data are taken in every clock pulse and shifted through the memory until they "disappear" at the other end. To visualize the data of interest, we need to be able to freeze the data flow at a required spot. This is done with the aid of the trigger function, a very important part of every logic analyzer.

**TRIGGER FUNCTION** The trigger function tells the logic analyzer to stop data acquisition and to freeze the data present in the memory at that instant. At the appropriate moment we take the data stored, format them, and transfer them to the display—the third main function of the logic analyzer.





**Fig. 19.13** (a) Use of direct leads for data sampling from a microcomputer system. (b) Actual hookup to a logic analyzer.

**DATA ACQUISITION** The sampled data should, of course, be acquired correctly from the system under test. That is done via probes, with color-coded leads being used to connect the probes to the system under test.

Depending on the type of analyzer, we also can use *individual* probes for connecting each data input to the circuit of interest, e.g., for sampling the data present on the data bus and address bus of a microcomputer system as shown in Fig. 19.13a. Figure 19.13b shows an actual hookup to a logic analyzer.

**DISPLAY** The data are stored in the memory in the normal binary code; however, the display can be formatted in various different ways, as shown in Fig. 19.14. In most logic analyzers the data displayed are preceded by a line number (related to the trigger word) for the sake of convenience. We now discuss the various possible display formats.

**Time format** One possibility is to convert the 1s and 0s of the data stored in the memory to pulses; then the display has the form shown in Fig. 19.14a. Up to 16 waveforms can be displayed, since we have 16 inputs in this example. This display looks very much like that of a normal multitrace oscilloscope; however, it is obtained in quite a different way (as a sampled pseudo-timing waveform).

The pulses for a given data word are vertically under one another; from left to right, first

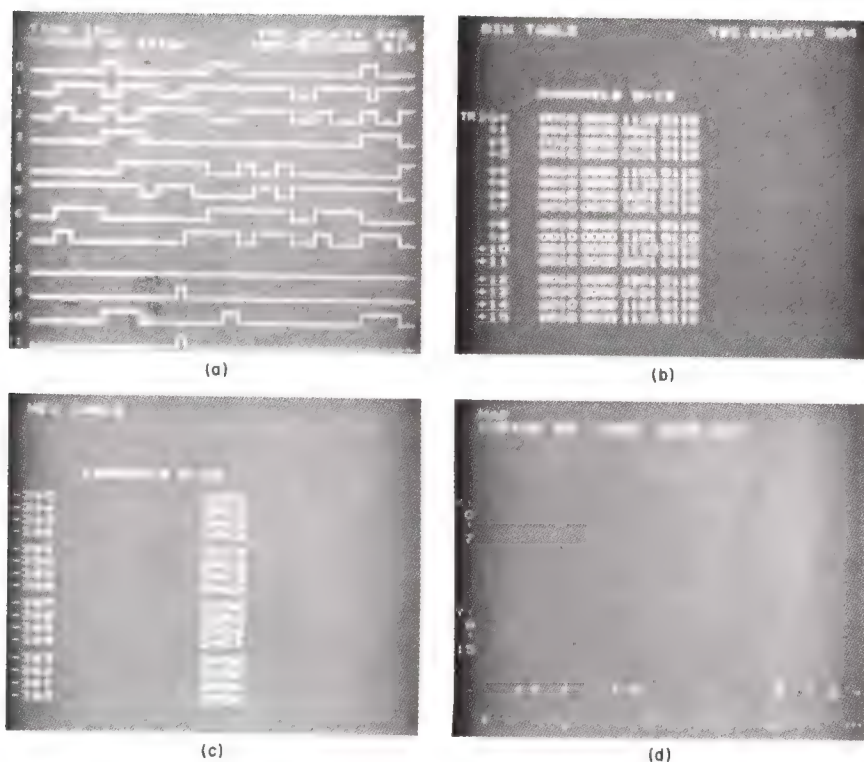


Fig. 19.14 Display formats.

we get all the bits for word number 0, then those for word number 1, word number 2, and so on.

Since the memory capacity is normally larger than the display capacity, the display shown on the screen represents a selection from the captured data in our memory, as seen through a "display window," which in principle can be given any desired width by the designer of the logic analyzer. In our example, the display window is 16 data words wide. The time format is used mainly for hardware analysis.

**Binary format** In the data domain, we often use other data representations. The binary format (Fig. 19.14b), for example, takes the same binary-coded data from the memory as for the time mode and represents them directly in 1s and 0s to give what is known as a *state table*. Here the individual data words are displayed horizontally, not vertically as in the time format. The display window is again 16 data words wide.

**Hexadecimal format** A more compact, and often more convenient, state table is the hexadecimal representation shown in Fig. 19.14c, which is simply a translation of the binary format of Fig. 19.14b to the hexadecimal notation (where the numbers from 0 to 15 are represented by 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, and F, successively). For example, D7 in hexadecimal = binary 11010111 = 215 in normal decimal notation.

Hexadecimal is widely used by people in the microcomputer world as well as by many minicomputer users—so why not use it in the analyzer too? A variant of this is the octal format, applied by some minicomputer users.

**Map format** A completely different display mode is the map format illustrated in Fig. 19.14d. This has many interesting applications as a kind of routing indicator for analyzing program flow. Once again, we take the same data from the analyzer memory, but convert them in another way. We come back to this map display later. The various display formats are generally pushbutton- or switch-selected in a logic analyzer.

## 19-10 Logic Analyzers, Logic Probes, and Signature Analysis

Summarizing our brief account of the operation of the logic analyzer, we see that this instrument takes in multichannel data (in our example, one new 16-bit data word every clock pulse) and stores the data in the memory. The trigger function stops data flow at the desired point. Then the data are displayed in a convenient way.

It is important to realize that the three main functions—acquisition, triggering, and display—are completely independent of one another.

### 19.6 MAIN FUNCTIONS OF A LOGIC ANALYZER

The three main functions of the logic analyzer, grouped around the memory where the sampled data are stored, are data acquisition, trigger function, and display. We now discuss each of the functions in greater detail.

#### DATA ACQUISITION

**The clock** A more detailed diagram of the data input (acquisition) unit is given in Fig. 19.15. We see here once again the memory, the data input and the *clock*, introduced above. The data input and the *clock* belong together: a new data word is picked up each time a clock pulse is given. We consider first the case where the clock belongs to the system under test—the physical source of input for the analyzer. This is denoted by “system clock” in Fig. 19.15. The data naturally run synchronously with this clock.

The system clock need not be the master clock of the system under test; indeed, this is seldom the case. It is often a derived clock or a control line (read-write or chip select line, for example). In any case, the system clock we select in our test object must be synchronous with the data we want to analyze. This type of operation is the *synchronous mode*.

We can use the analyzer's internal clock also. This is completely independent of the system clock and thus runs asynchronously. The internal clock is a free-running signal produced by the analyzer, having no time relation with the incoming data. We explain later the reasons for using this asynchronous mode.

**Data input** Before we can transfer data from the real world via the input to the analyzer's memory, we must choose the appropriate range of input levels by means of the threshold level selection (TLS) unit, shown in Fig. 19.15. The threshold level is determined by the logic family in the system under test (for example, TTL, ECL, or MOS).

Figure 19.16 shows the situation for a single data input line. The switching threshold for

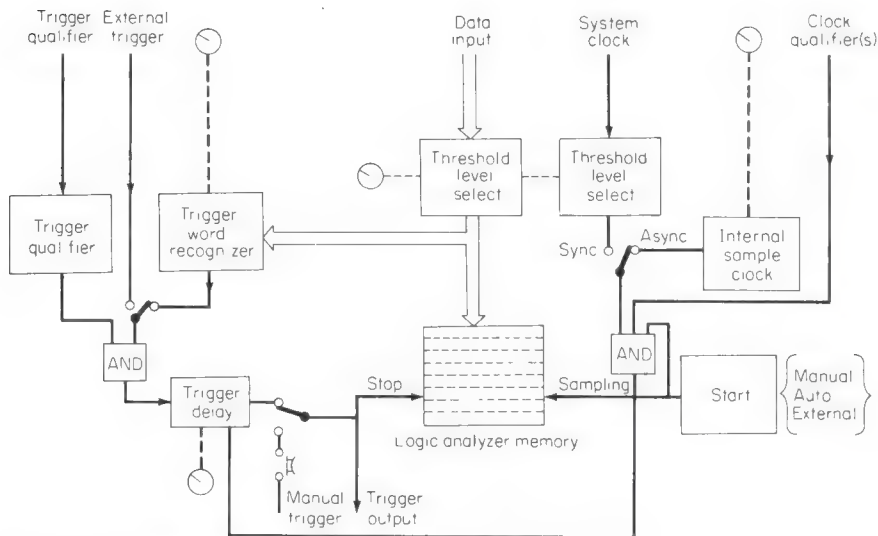


Fig. 19.15 Clocking data into the LA memory.

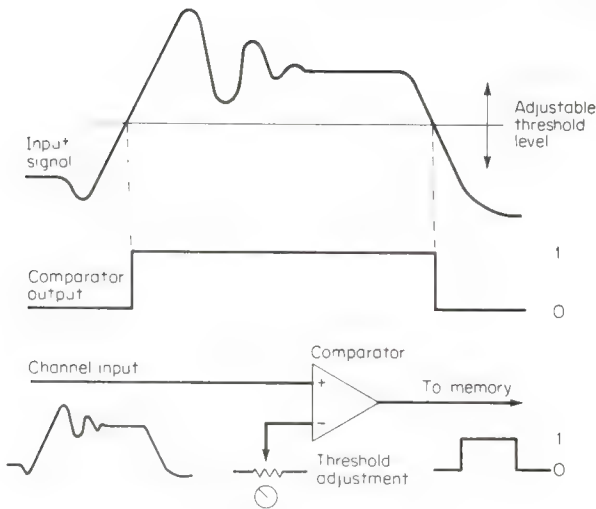


Fig. 19.16 Input with threshold detection.

measurements on TTL circuitry is about 1.5 V. If we set the TLS unit to this value, every incoming voltage higher than 1.5 V is regarded as being a logic 1 and every input voltage below 1.5 V as a logic 0.

The comparator output signal indicated in Fig. 19.16 is sent to the memory for sampling.

**Setup and hold times** Figure 19.17 shows the timing diagram for one input channel, with data at the input and a clock pulse. Now, it is not the clock pulse as a whole but rather its leading edge that determines the moment of data sampling.

If we take a positive-going clock pulse, the positive-going *edge* is important. It is also possible to use the negative-going edge of a negative clock pulse, as shown in Fig. 19.17. The choice between these two types of clock pulses is user-selectable.

Now, the data must be present for a certain time (called the *setup time*) before the clock edge appears, to make sure that the data at the entry to the analyzer memory are completely stable when the clock edge does appear. The setup time is of the order of a few nanoseconds. The maximum permissible clock rate is also an important parameter, and it is related to the setup time. It is generally between 10 and 100 MHz (corresponding to clock pulse rates between 100 and 10 ns).

The hold time is another important parameter in this connection. The data should remain stable for a short time after the sampling edge, to ensure correct data sampling. This hold time is generally very near to zero.

Another important factor is the duration of the clock pulse itself, even though data sam-

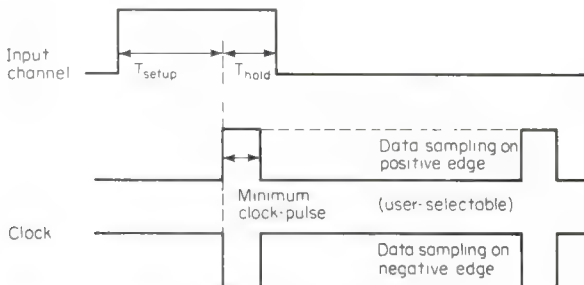
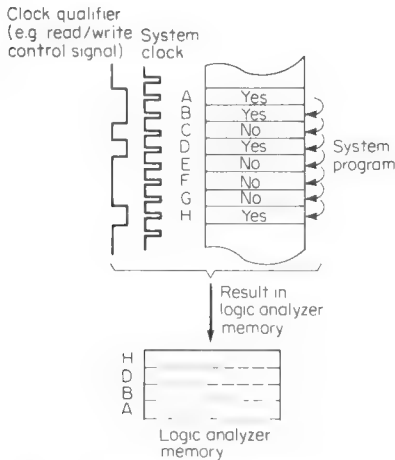


Fig. 19.17 Definition of setup and hold times.

pling is basically controlled by the leading edge of the clock pulse. This should generally be of the order of a couple of nanoseconds.

**Clock qualifiers** The synchronous mode is used mainly in the data domain. For selection of data in a data stream, wide use is made of the clock qualifier. The word *qualifier* sometimes gives rise to confusion. There is also another qualifier, called the *trigger qualifier*, which has a completely different function and is discussed later.

A clock qualifier is used for selective sampling of the data words wanted from the data stream. Application of the clock qualifier comes down to gating the clock with an additional condition. This is illustrated in Fig. 19.18,



**Fig. 19.18** Clock and clock qualifiers (selective sampling).

so on. The overall result is that we store the four data words marked "yes" out of the total of eight data words in the memory.

In another application, we might study the write operation. For this purpose the logic assignment of the clock qualifier is reversed, and we take the high level as false and the low level as true.

One main advantage of the clock qualifier is that it permits us to make more efficient use of the memory capacity of the logic analyzer. A disadvantage is that the time relation in our captured data is lost, since only the selected data words are stored. However, data acquisition is still synchronous with the enabled clock pulses. As long as we are not interested in time-related data sequences, this will not matter.

In the data domain, we are interested in related sequences of data, so a clock qualifier is of great importance. Use of a number of clock qualifiers permits selection of the data words in even greater detail or depth.

**Start function (analyzer begins to gather data)** One function not mentioned so far is the *start* function used to initiate data acquisition. This may be manual—we press the button and sampling starts. It also may be automatic with starts at regular intervals of, say, 1 s for automatic pseudo-real-time measurements in the data stream or via the external input with an electric signal (e.g., a TTL signal).

**TRIGGER FUNCTION** As long as data acquisition is not stopped, the data flow through the memory continues and no data are displayed. Data acquisition is stopped with the aid of the trigger function, which is a very important part of every logic analyzer. It permits us to capture only those blocks of data from the data stream in which we are interested. Unlike the trigger in an oscilloscope (which *starts* the time base), the trigger in a logic analyzer *stops* data acquisition. We discuss the trigger function with reference to Fig. 19.19, where the crosshatching represents the trigger section. Triggering can be realized in various ways: externally, internally, or manually.



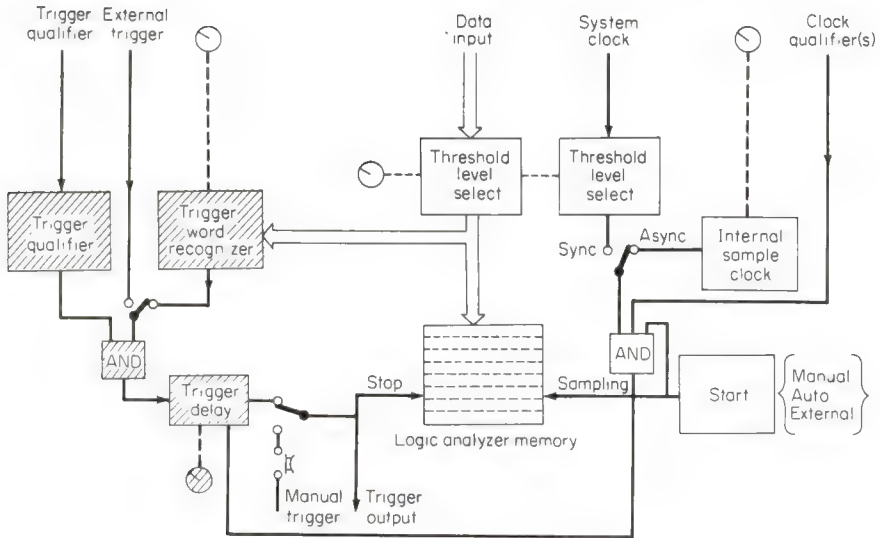


Fig. 19.19 Trigger function.

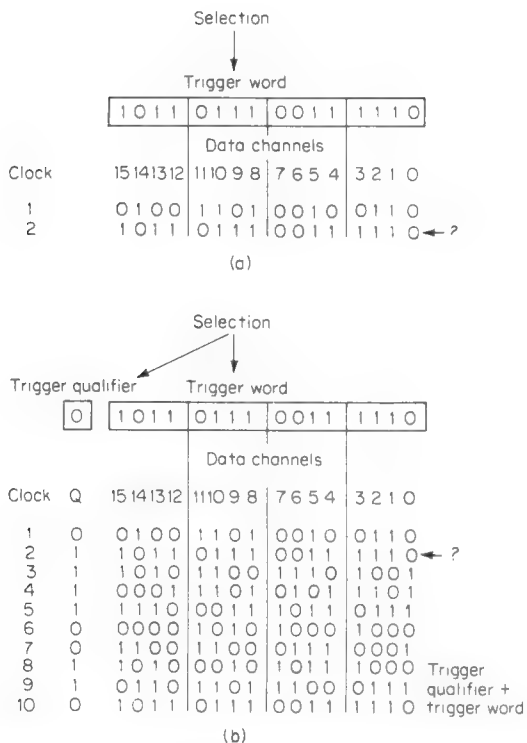
**Manual triggering** Here we just press a button to stop further data acquisition. This manual trigger is, in fact, a manual stop. It is not normally used, but we see later that the other trigger facilities do not always give us correct triggering under all circumstances. For example, if the measuring conditions are such that data acquisition keeps on continuously and the "not triggered" signal appears on the screen, then we can use the manual stop to freeze the data passing by at that moment, to get some idea of what is going on. We could compare this with a beam finder on an oscilloscope, which also gives an indication of what is the matter with the measurement.

**Internal triggering** Internal triggering (also called *combinatorial* triggering) is performed with the aid of the *trigger word recognizer*. This is the most important trigger facility, and it is used in the majority of measurements. Here we select the trigger word with the operator-controlled trigger word selector, and we stop the data flow as soon as the last data word sampled from the data stream coincides with the trigger word. This trigger word can be selected with mechanical switches, but there are other means, depending on the analyzer. Philips PM 3500 logic analyzer uses three-way switches, which can be set to 1, 0, or "don't care" (immaterial) at each position. This enables us to select one unique trigger word from  $2^{16}$  possibilities (we have 16 inputs). In Fig. 19.20a, we see such a selected trigger word, with a match between the trigger word and incoming data at the second captured data word. Only then will the trigger circuit generate the stop pulse.

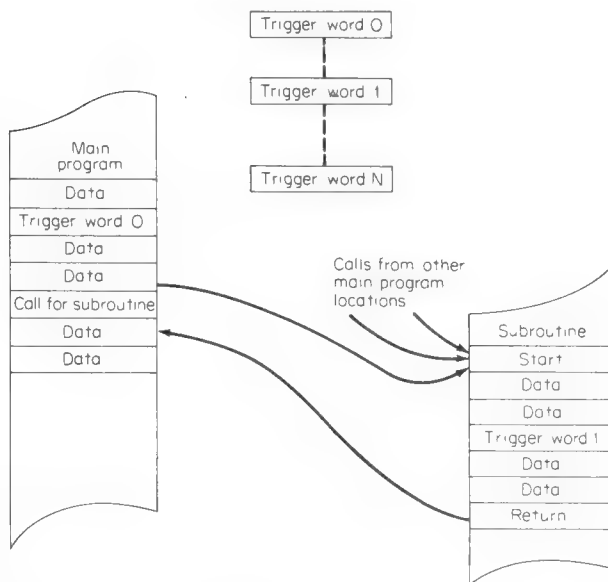
Analyzers now use what is known as *sequential*, or *nested*, triggering. This means that not one trigger word is selected but several are one after another. All trigger words should be matched in a selected sequence before the stop pulse can be given. This is illustrated in Fig. 19.21, which shows part of a program where both trigger words 0 and 1 have to be matched before a stop pulse is given and the data can be displayed. This permits data analysis before and during this conditional call as well as the subroutine. If the subroutine were called from another place in the main program, we would not get a stop pulse.

Further additional conditions, e.g., that a certain time should elapse or a certain event should occur between trigger words 0 and 1, another between trigger words 1 and 2, and so on, are sometimes built in. The *trigger qualifier* is an extra input, an extra qualification for triggering. If we combine this trigger qualifier with the trigger word recognizer, we could also speak of a "17th input channel." Although there is no storage or display of data for this channel, it is taken into account for triggering. Not only should the trigger word be matched, but also the trigger qualifier condition should be fulfilled before a trigger pulse is given to stop data acquisition. We see in Fig. 19.20b that both these conditions are not





**Fig. 19.20** (a) Trigger word. (b) Trigger word and trigger qualifier.



**Fig. 19.21** Nested triggering.

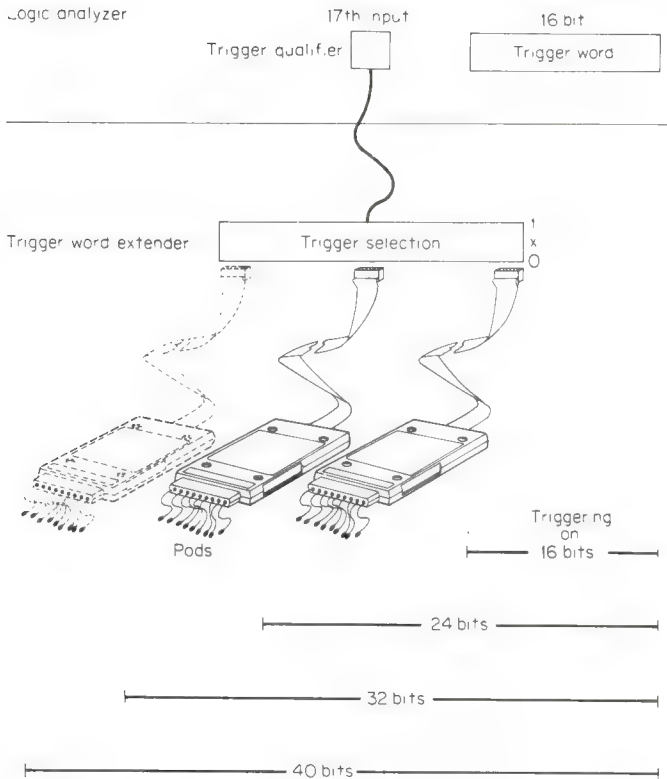


Fig. 19.22 Trigger word expansion.

fulfilled by the second data word in the table, so data acquisition will not be stopped here after all; we have to wait for the last data word in the table before this occurs. This 17th channel can be enlarged in multiples of 8 by means of trigger pods, as illustrated in Fig. 19.22. This enables us to select one unique trigger word out of a much larger data word. Switches are provided on the trigger probes to select a 0, a 1, or a "don't care" for each additional trigger input. This is called a *trigger word expander*.

**External triggering** The logic analyzer also can be triggered externally via an input which accepts an electric signal from such sources as an interrupt line, a timer, or a specific control line of the system under test. In all three trigger modes—external, internal, and manual—the trigger function works independently and can be used in exactly the same way in synchronous and asynchronous sampling.

**Trigger delay** Trigger delay allows us to select from the data stream a data block that comes a preselected number of clock pulses after the selected trigger word. This is a delay by a specific number of samples, not by a specific time. The delay can be user-selected to any integer number between 0 and 10 000, say. It also can be switched off, which is functionally the same as zero delay. If a certain delay is selected, for example, 63, an internal down counter in the delay circuit will count down, using the same clock pulse as for clocking new data in the analyzer memory. We would like to explain the use of the trigger delay with reference to the simple program of Fig. 19.23. On receipt of every sample pulse, one new data word is captured and is shifted through the analyzer memory in the data stream.

If no trigger delay is selected (Fig. 19.23a), then as soon as the trigger word recognizer finds a match between data and selected trigger word, it produces an immediate stop. The data are displayed, and new data are no longer sampled. The last data word stored is our selected trigger word. We have captured historical data, i.e., data occurring *before* the



words to be sampled before the stop pulse is given. Since our memory capacity is eight in this example, the trigger word itself is not displayed (it has been shifted through the analyzer memory long ago), though its value is, of course, still set on the trigger word selector. The trigger delay enables us to cover a range of historical data as large as the memory capacity and of future data as large as the delay (1000 data words in this example); thus we effectively stretch the memory.

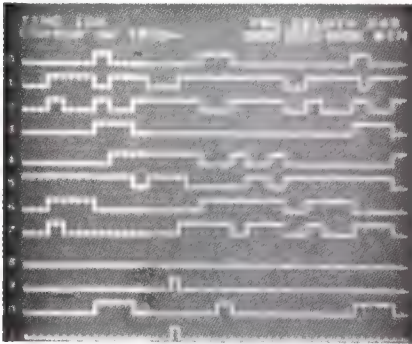
The trigger delay function can be regarded as a means of trigger positioning at an appropriate point in our data stream. The unique trigger word provides a reference point, and the trigger delay positions the required block of data from the data stream relative to this trigger word.

The stop pulse can be used for other purposes apart from stopping data acquisition. For example, we can trigger another instrument such as a counter or an oscilloscope with it, via the trigger output. This instrument can then carry out a supplementary measurement, related to a specific point in the data stream.

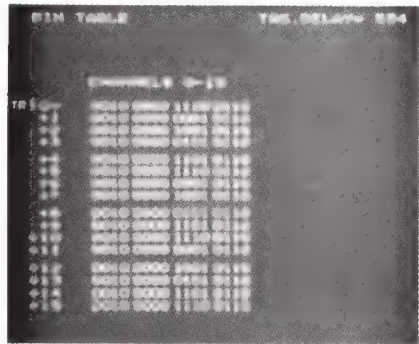
**DISPLAY** The display function links the captured data in the analyzer memory to the user's interpretation of the data display. It takes the data as they are stored and presents them to the user in a convenient format.

Let us start by recapitulating some of the facts about the display mentioned previously (see Fig. 19.24). The *binary* representation (Fig. 19.24b) may be regarded as the most straightforward. It provides a state table giving the data in 0 and 1 form for all 16 inputs.

The *hexadecimal* representation (Fig. 19.24c) also provides a state table, but here the binary data are combined in groups of 4 bits, each of which is replaced by a hexadecimal character. A 16-bit word length thus corresponds to four hexadecimal characters. The octal state table compresses the data in a similar way, by groups of 3 bits. Other formats such as ASCII and DECI as well as binary and hexadecimal are available.



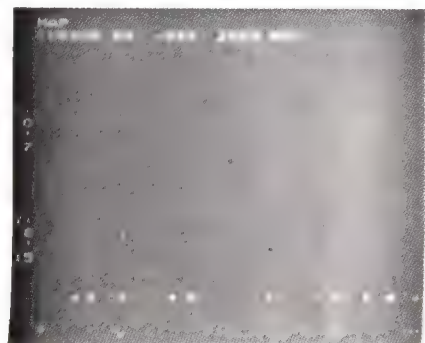
(a)



(b)



(c)



(d)

**Fig. 19.24** Display formats (see also Fig. 19.14).

Another useful feature in formatting the display is the possibility of decoding the words in the data stream into mnemonics representing disassembled microprocessor instructions, e.g., for the 8080/85 family,

-2	7E	MOV AM
-1	80	ADD B
0	C9	Return

or a combination of IEC bus commands and data:

-3	LAD 4
-2	DATA
-1	DATA
0	GET
1	TAD 7

With the aid of the *cursor control*, we select the block of stored data we want to display. In Fig. 19.25, this block contains 16 lines of 16-bit data words out of a total memory capacity of  $504 \times 16$  bits. This cursor-controlled part is called the *display window*. By turning the cursor control we can, as it were, move the display window past the stored data until we bring to light the data block of interest. Additional control is provided by the *blanking* facility. If we are interested in only a limited number of channels (for example, only 5 bits), we can blank the other ones. Remember that this is only a display control; for example, the trigger word still has to be set up in full on the trigger word selector, and the full data word remains in the logic analyzer memory.

In the *compare mode*, we compare the new data sampled in a given measurement with reference data already stored in an additional memory. Each data bit that is not equal to the corresponding reference bit is intensified in the display. Every bit of the entire contents of the memory is compared with the corresponding reference bit, but only a selected part is shown in the display window. Additional information is, therefore, given above the display, in the form of a message *equal* (the data in the memory are completely identical with the reference data) or *unequal* (one or more bits out of the total memory contents are different). We can then move the display window up and down to trace all the different—intensified—bits.

In the *time format* (Fig. 19.24a), the display waveforms look just like those from oscilloscopes. However, as we mentioned above, these waveforms are not real-time as in an oscilloscope, but just represent the stored data from the analyzer memory reconfigured into a pseudo-time display. The time format is used mostly for measurements in the hardware environment.

The time mode can be used to display the entire contents of the memory (by setting the magnification control to " $\times 1$ "). A part of this display (the *cursor block*) is intensified and can be moved along the memory with the aid of the cursor control. The first data word of this cursor block is also displayed, together with the line number relative to the trigger word and the contents of this data word in binary, hexadecimal or octal notation. The same cursor block will stay on display when we select only a part of the memory contents (magnification  $\times 10$ ) or an even smaller part (magnification  $\times 25$ ). Blanking also is possible here; the remaining channels will be regrouped to fill the entire display for user convenience.

A completely different type of display formatting is provided by the *map mode* (see Fig.

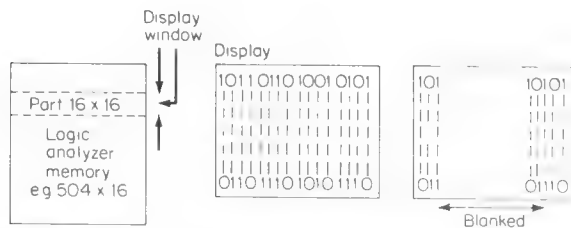


Fig. 19.25 Display window and blanking.

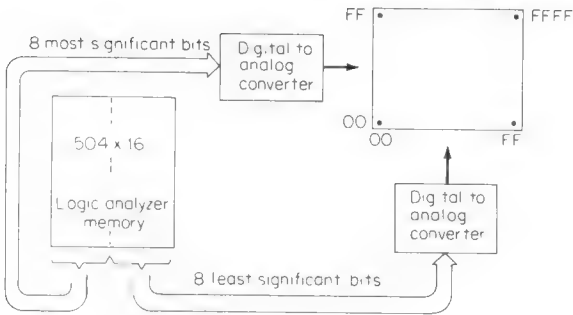


Fig. 19.26 Map display.

19.26). This gives us an overview of all captured data words (504 in the present examples) in what we might call a geographic way. The map format is used mostly in data-domain measurements. For this purpose, each data word in the analyzer memory is split into two halves, the least significant byte and the most significant byte. The contents of each byte is processed by a separate digital-analog converter (DAC), to yield a deflection of the spot in the horizontal direction for the least significant byte and in the vertical direction for the most significant byte. Each data word is thus converted to a point on the display screen, characterized by its own  $xy$  coordinates.

The values of the 2 bytes determine the unique position of the spot for each different data word. If we have 504 locations in the memory, 504 different spots can be displayed at the same time. If we now take the address lines of the system under test as data input for the analyzer (see Fig. 19.27), we can display the flow of a captured part of the program in map mode. Remember, the value of the program counter, present on the address bus in some form, always indicates the *next* program step.

The example of Fig. 19.28 shows the map display of a program structure with several subroutines stored at different address locations. The addresses of this program are sampled and stored in the analyzer's memory. The specific pattern in which the map-mode dots appear on the display characterizes our captured program.

The first addresses sampled from the main program appear here as the dots in the bottom straight line of Fig. 19.28, where successive dots represent successive addresses. At a certain point in the program there is a call for the first subroutine; the start address for this subroutine is thus the next address sampled. This is clearly indicated on the display, since the next spot is located at a different place (1 in Fig. 19.28).

At the end of this subroutine we return to the main program and continue the straight line at the bottom of the display. At the second call for this first subroutine, the next address is again the start address of the subroutine. The spots on this subroutine are more intensified

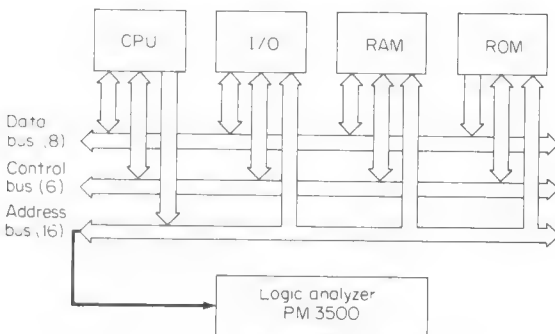


Fig. 19.27 Application of map display mode to a microcomputer system.



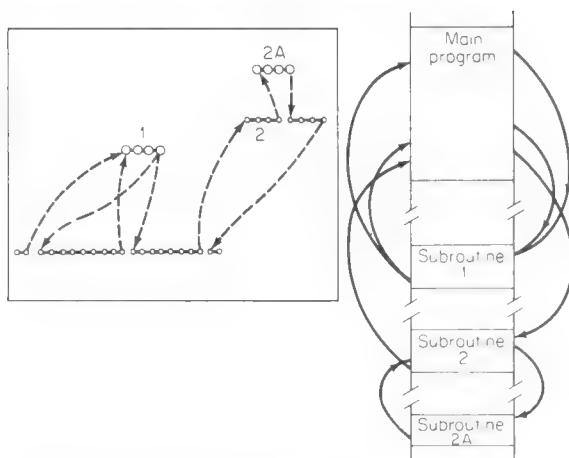


Fig. 19.28 Map display of program flow.

than the ones from the main program because the subroutine is executed twice, the main program once. This gives an easy way of recognizing parts of the program that are executed more often than others. The remaining part of Fig. 19.28 can be understood in much the same way.

**SYNCHRONOUS AND ASYNCHRONOUS SAMPLING** Let us now return to the data acquisition process. We can take a system clock as sample pulse, synchronous with the sampled data for data-domain measurements. For the hardware environment, however, this does not always satisfy the measuring requirements. There are often conflicting timing relationships or short, unwanted signals such as spikes and glitches which could disturb the correct functioning of a digital system. We find that not all the data are captured in the synchronous mode (see Fig. 19.29). The solution is to select a faster clock. If a clock with higher repetition rate is not available in the system under test, we take the logic analyzer's internal clock with (normally) a rate that is roughly 5 to 10 times faster. The repetition time of this clock can be user-selected between, say, 10 ns and 20 ms, but this clock is completely independent of, or asynchronous with, the data.

Figure 19.30 shows an example of asynchronous sampling of the data, at a much higher clock rate. For the sake of convenience, Philips' PM 3500 logic analyzer shows the clock ticks, which are given for easy time reference in the time display mode.

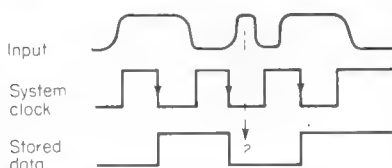


Fig. 19.29 Synchronous data sampling.

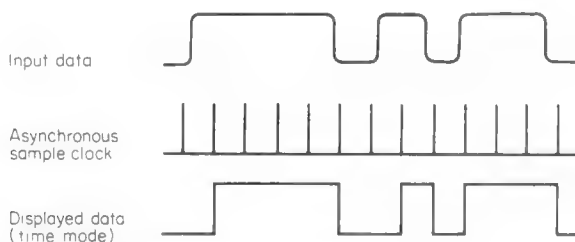


Fig. 19.30 Asynchronous data sampling.

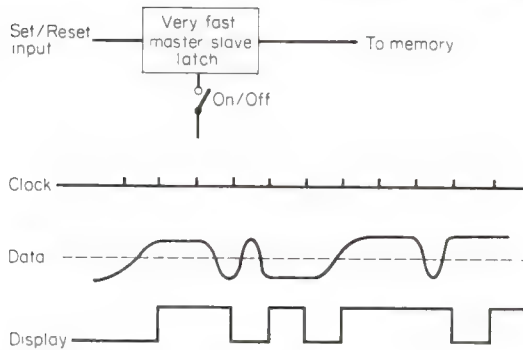


Fig. 19.31 Glitch-capturing mode.

**GLITCH-CAPTURING MODE** So far, we have considered only data that are taken in at clock-pulse edges (synchronously or asynchronously). This is called the *sample mode*. It is not possible to see data transitions in between clock pulses in the sample mode. However, this is possible in the *latch* or the *glitch-capturing mode* (Fig. 19.31). In this mode an additional, very fast master-slave latch located immediately after each input is switched on. We also could say that an extra memory capacity of one 16-bit data word is used to capture signal transitions in between clock pulses.

Each initial transition of the data signal from its original state to the opposite one (1 to 0, or 0 to 1) will be stored in the fast latch, and this information will be transferred to the normal working memory. Even glitches down to 3 ns can be captured in this mode.

## 19.7 APPLICATIONS OF THE LOGIC ANALYZER

**EXERCISES IN MONITORING DATA TRANSPORT VIA BUS LINES** In this section we give some elementary exercises that could be used to familiarize the trainee electronics engineer (or anyone new to the logic analyzer) with the working of this instrument.

In every microprocessor system, the data are transferred over the buses to various peripherals such as memories and I/O devices. A logic analyzer must be capable of acquiring and monitoring this information in the most convenient way. We give below simple exercises to illustrate some ways in which the logic analyzer functions to this end.

All our examples refer to the basic microprocessor system of Fig. 19.32, consisting of a 8085 microprocessor, 8355 PROM/IO (programmable read-only memory with associated input-output devices), 8155 RAM/IO (random-access memory with associated input-output devices), and 8205 address decoder.

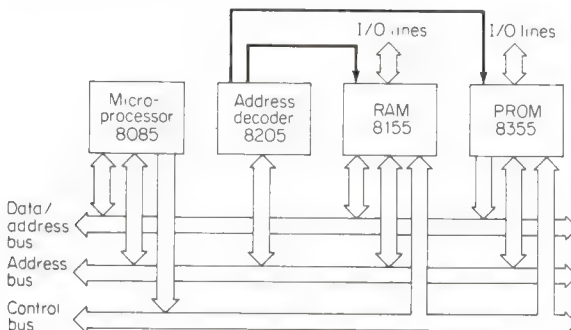


Fig. 19.32 Block diagram of microcomputer setup used as a basis of the exercises in this section.

TABLE 19.1 Program Used in Exercises

Address	User label	Instruction mnemonic	Operand or address	Description
0575	XMPL4	LXI	H, 2000H	Load 2000H in register pair HL
0578		MVI	A, 0DH	Set port A to write, port B to read
057A		OUT	20H	Write to RAM
057C	XM41:	INR	B	Increment counter (add 1 to contents of B register)
057D		MVI	A, 0CH	Move 0CH (hexadecimal notation for decimal 12, or digital 0000 1100) to accumulator
057F		SIM		Mask interrupt 7.5
0580		OUT	23H	Write 0C to port A
0582		MOV	A, B	Move contents of B register into A
0583		OUT	21H	Write to port A
0585		IN	22H	Read port B
0587		MVI	M, 0FFH	Write FF (hexadecimal notation for decimal 255, or digital 1111 1111) to address 2000
0589		JMP	057C	Go and get next number

We assume that the program shown in Table 19.1 has been loaded into the PROM. It would take us too long to explain this program completely to someone unfamiliar with the workings of a microcomputer, but the following information may prove useful.

**Address.** Addresses are the locations in the memory in which the data and instructions are written (have been stored). The addresses are given in hexadecimal notation, in which the digits 0 to 9 have their normal decimal significance and the letters A to F are used to represent decimal 10 to 15. Thus, e.g., the address 057F corresponds to  $5 \times 16^2 + 7 \times 16^1 + 15 \times 16^0 = 1280 + 112 + 15 = 1407$  in decimal notation. It is customary to write hexadecimal numbers with a subscript H (for example, 2000<sub>H</sub>) to distinguish them from decimal numbers of the same form.

**User Label.** These codes have the following significance here: XMPL4 = example 4 and XM41 = first label in example 4.

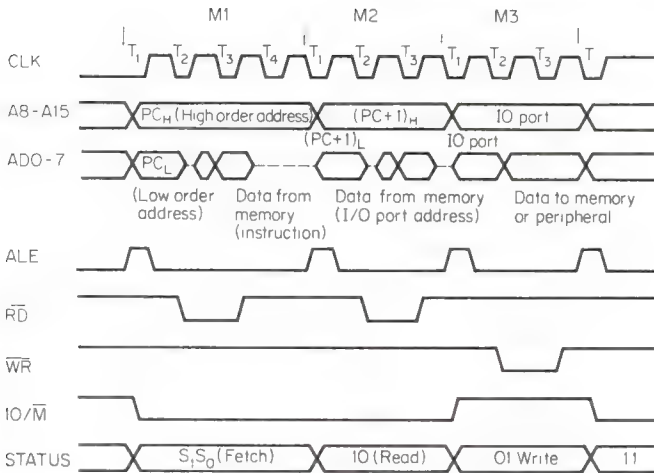
**Instruction Mnemonic.** These mnemonics (also known as *op codes*, short for “operational codes”) are more or less comprehensible representations of the relevant instructions taken from the instruction set of the 8085 microprocessor. The mnemonics used in this program have the following significance:

- LXI Load immediate register pair
- MVI Move immediate register
- IN Input
- OUT Output
- INR Increment register (add 1 to current contents of register)
- SIM Set interrupt mask
- JMP Jump unconditional

In fact, each of these mnemonics is stored in the PROM in the form of an 8-bit “machine code” which would be very difficult for the programmer to work with directly. The correspondence between the mnemonics and the machine codes comprising the instruction set may be found in the data sheet for the microprocessor in question. For example, the mnemonic SIM represents the machine code 00110000 (which corresponds to 030 in hexadecimal notation, or decimal 48 in the instruction set for the 8085).

**Operand.** Operands are also stored in the PROM in the form of 8-bit bytes. Each mnemonic may be followed by no data or address bytes, one data or address byte (if there is no comma in the entry in this column) or two such bytes (if there is a comma in the entry).

**Description.** Assembler programs used to translate from machine language to higher-level languages such as the instruction mnemonics given here include routines for printing out very brief descriptions of the various instructions, in order to guide the electronics engineer working on the development of a given microcomputer system. However, the descriptions given by the assembler are so cryptic as to be hardly intelligible to any but the insider; we have, therefore, expanded the description somewhat here.



**Fig. 19.33** Basic timing diagram of 8085 microprocessor.

The overall effect of the program listed above is to ensure that the sequence of (decimal) digits 0, 1, 2, 3, 4, etc. is written to input-output (I/O) port A of the 8155 RAM, and port B is read immediately afterward. Also, each time one digit from the above sequence is handled by ports A and B, the hexadecimal combination 0C is written to port C and the combination FF to memory location 2000H. We thus get the data sequence 00, FF, 0C, 01, FF, 0C, 02, FF, 0C, 03, FF, 0C, etc. passing along the data bus lines.

Figure 19.33 shows the basic timing diagram of the 8085 microprocessor. A timing diagram is a fundamental tool for logic analysis, for it gives all sorts of information such as types of data, control and status information, parameter setting, clock qualifiers, etc. at the appropriate points with the PROM addresses for the RAM (2000H) and I/O (2121, be used for all these exercises.

**Exercise 1** First we want to check all the addresses used in this program.

#### Connections

Clock input of logic analyzer to ALE (address latch enable) lines

Eight data channels of PM3540 to most significant address bus lines A8-A15

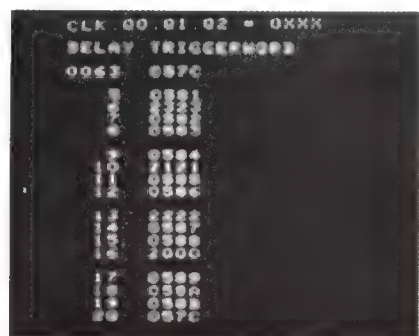
Eight data channels of PM3540 to least significant address bus lines AD0-AD7

#### Settings (see heading of Fig. 19.34)

It may be seen from the timing diagram that all addresses are stable on the buses during the negative-going edge of the ALE pulse.



(a)



(b)

**Fig. 19.34** (a) Checking the program addresses. (b) Display is "rolled up" five addresses in order to see entire program.

## 19-24 Logic Analyzers, Logic Probes, and Signature Analysis

A falling edge ( $\text{CLK} = 0$ ) has, therefore, been chosen as clock parameter for the logic analyzer.

The trigger word can be chosen at any convenient point of the program flow; we took 057C here.

Post-triggering can be chosen in order to capture *all* the information after the trigger word; a delay of 63 is chosen to get all the information plus trigger word in the logic analyzer's 64-word memory.

**Results.** In Fig. 19.34a we see a photograph of the display of the PM3540, showing all the addresses from 057C to 0588 at which the various instructions are written, interspersed at the appropriate points with the PROM addresses for the RAM (2000H) and I/O (2121, 2323, etc.).

By rolling or paging through the memory contents the whole program can be displayed; Fig. 19.34b shows the display rolled up to reveal the addresses from 0581 to 0588.

**Exercise 2: Checking a selected part of the program flow (addresses)** An address decoder decodes addresses so as to activate enable lines for the various devices connected to the system, thus selecting which chip is to be used in the system at a given time (here either the PROM or the RAM). To select the data, we use the logic analyzer's clock qualifier, which gates the clock pulses.

### Connections

Clock (CLK) input of logic analyzer to ALE line

Clock qualifier  $Q_0$  to CS (or CE) of PROM

Data channels as in Exercise 1

### Settings

CLK = 0, that is, negative-going edge of ALE is active

$Q_0 = 0$ , since CS signal is active when low

Trigger word: any word in 05xx series or xxxx can be chosen (here we choose 057C again)

Delay: 63

**Results.** As may be seen from Fig. 19.35, only the addresses of the program as such (057C–058B) are displayed now; those of RAM (2000) and I/O ports (2121, 2323) are skipped.

**Exercise 3: Analyzing the data flow** If we want to see the data, we cannot use the ALE signal to clock the logic analyzer. As may be seen from the timing diagram of Fig. 19.33, either the  $\overline{\text{WR}}$  (write) or  $\overline{\text{RD}}$  (read) pulse can, in general, be used as clock to pick out the data flow. In this case, we are interested in seeing the data written to the memory and the I/O ports; the write pulse is thus the appropriate choice here.

### Connections

Clock input of logic analyzer to  $\overline{\text{WR}}$  line

Data channels as in Exercise 1

### Settings

CLK = 1, since  $\overline{\text{WR}}$  is active on positive-going edge (see timing diagram of Fig. 19.33)

Since the 8085 is an 8-bit multiplexed microprocessor, only bus lines AD0–AD7 transfer data. The most significant half of the address, which is also clocked into the analyzer, thus is not relevant information for this exercise and can be blanked. Hence, we choose the trigger word xx00 (or ??00 in hexadecimal notation, where ? replaces the binary "x" or "don't care" sign).

Delay: 63

**Results** The data flow (00, FF, 0C, 01, etc. . .) in the programmed sequence is displayed in Fig. 19.36.

**Exercise 4: Selecting I/O data only** As the timing diagram of Fig. 19.33 indicates, the signal  $\text{IO}/\overline{\text{M}}$  determines whether data are transferred to I/O ports or to the memory. We can select, e.g., the data to I/O ports by feeding this  $\text{IO}/\overline{\text{M}}$  signal to the clock qualifier input of the logic analyzer.

### Connections

Clock input of logic analyzer to  $\overline{\text{WR}}$

$Q_0$  to  $\text{IO}/\overline{\text{M}}$

Data channels as Exercise 3

### Settings

CLK = 1





**Fig. 19.35** Program address listing that skips the addresses of the RAM and I/O ports.



**Fig. 19.36** Analyzing data flow.

$Q_0 = 1$ , because data are sent to I/O ports when  $\text{IO}/\overline{\text{M}}$  is high (see timing diagram of Fig. 19.33)

Trigger word: xx00

Delay: 63

**Results.** Only the information passed to the I/O lines is displayed, as shown in Fig. 19.37. If we had selected  $Q_0 = 0$ , then only the data written to the memory would be shown, i.e., an uninterrupted sequence of FF, FF, FF, FF, . . .

**Exercise 5: Selecting only data written to port A** If we call up the most significant half of the addresses, other settings and connections being as for Exercise 4, we get the display shown in Fig. 19.38.

It can be seen that the most significant half of the address of port A is  $21_{\text{H}}$ , and that of port C is  $23_{\text{H}}$ . In binary terms this means that these ports differ only in the second bit of their address, so this bit gives the vital information on what port is called. A second clock qualifier input receiving this bit can thus be used to select data only from port A—or only from port C.

#### Connections

Clock of LA to  $\overline{\text{WR}}$

$Q_0$  to  $\text{IO}/\overline{\text{M}}$

$Q_1$  to A8

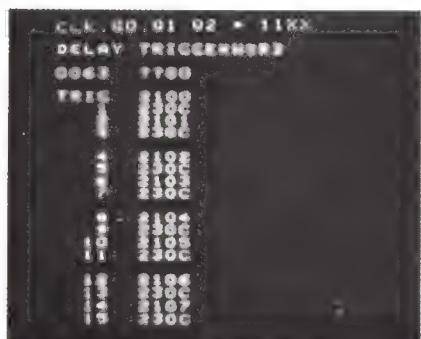
Data channels as in Exercise 4

#### Settings

CLK = 1



**Fig. 19.37** Data flow through the I/O ports only.



**Fig. 19.38** As in Fig. 19.37, but with the most significant half of the addresses displayed (recalled).





**Fig. 19.39** Data flow to port A, selected on basis discussed in caption to Fig. 19.38.

$Q_0 = 1$

$Q_1 = 0$  for selecting port A

Trigger word: xx00

Delay: 63

**Results.** Figure 19.39 shows only the data written to port A. If we had selected  $Q_1 = 1$ , only the data written to port C (i.e., an uninterrupted sequence of 0C, 0C, 0C, etc.) would have been displayed. Note that we would have needed a new trigger word—xx0C—for this purpose.

## 19.8 GLOSSARY OF LOGIC ANALYZER TERMS

The following “dictionary” of terms commonly associated with the logic analyzer (generally abbreviated LA below) has been split into a number of sections dealing with triggering, sampling, display functions, memory considerations, and general topics. A final section dealing with common microprocessor terminology has been included as general information.

### Triggering

**Center Triggering.** When the memory contains data leading up to and immediately following the trigger event.

**Delayed Trigger.** When the memory contains data that occur a present digital delay after the trigger event.

**Edge Triggering.** Trigger word recognition occurring on a clock edge which controls the sampling of an external input.

**External Trigger.** An additional input channel, usually without data storage, which acts as a single channel trigger.

**Glitch Triggering.** The ability of an analyzer to trigger on extremely short pulses which it would not normally sample into its memory.

**Manual Triggering.** User triggers the logic analyzer by pushing a button.

**Pass Counter.** A selectable counter used in sequential triggering. Its action is to delay the arming of the second word until a preset number of passes of the first have occurred.

**Post-triggering.** When the memory contains the trigger event and the data immediately following it.

**Pretrigger.** When the memory contains the data that occurred up to and including the trigger event.

**Sequential or Nested Triggering.** Triggering of the LA which only takes place in a certain sequential logic flow; for example, word A must be followed by word B before triggering occurs. Two or more trigger words must be selectable in the logic analyzer. It is used to isolate a particular conditional entry to a subroutine at different points.

**Trigger Qualifier.** An additional input that must be “true” before the word recognizer can recognize the trigger word.

**Trigger Word.** User-selected logic pattern on which the analyzer triggers.

**Word Recognizer.** The part of the LA that identifies the trigger word in the logic flow.

## Sampling

*Asynchronous Sampling.* The sampling rate of the logic analyzer is controlled by its own internal high-speed clock.

*Clock Qualifier.* A logic state input that is ANDed with the sampling clock edge. Only when this state input is true at the moment of the active clock edge does sampling occur.

*Clock Width.* Minimum clock high time or low time for sampling.

*Dual Clocking.* System enabling a logic analyzer to sample information from two clock edges (usually two separate systems in one box). Application area in multiplexed bus microprocessors where one clock determines the address and a second identifies the data at that address.

*Hold Time.* Time during which data must be stable after the active clock edge.

*Latching or Capture or Glitch-Detector Mode.* Logic analyzer operating mode based on use of fast input circuits detecting the presence of transient pulses that would not be sampled by the clock edge.

*Sample Interval.* The time between samples (= reciprocal of the speed).

*Setup Time.* Time during which data must be stable prior to the active clock edge.

*State Analyzers.* Those analyzers that operate with an external clock provided by the system under test.

*Synchronous or State Sampling.* The sampling rate of the logic analyzer is controlled by a selected edge in the system being monitored.

*Timing Analyzers.* Those analyzers that have the ability to operate asynchronously with the system.

*Timing Analyzer Speed.* The repetition rate of the sampling edge—not the bandwidth.

## Display

*ASCII (American Standard Code for Information Interchange), Equivalent to the ISO 7-Bit Code) Mode.* A display mode which converts a seven-digit binary value to alphanumeric characters. See Table 19.2.

*BIN (Binary Display Mode).* Each channel is shown as series of 1s and 0s.

*Blanking.* The removal of unwanted channels from the display.

*Data Field.* That area of the display which details the captured data.

*Display Window.* The area of the captured data that is displayed in the data field.

*Graph.* Display mode that plots the serial number of each captured item of data on the horizontal axis against the magnitude of the data at that location on the vertical axis.

*HEX (Hexadecimal Display Mode).* The binary value of four binary channels expressed as a hexadecimal value (0 to F).

*IEC 625 (IEEE 488, GPIB) Mode.* A display mode that decodes the binary information being transferred across the IEC interface bus.

*Inverted-Logic Display.* Logic high is a 0, and logic low is a 1. In this display mode, all values are inverted in state displays; but voltage levels remain true in the timing mode.

*Mapping.* A display mode that shows the full-memory depth in a single display. The high-order 8 bits control a digital-to-analog converter on the X axis, and the low 8 bits control a digital-to-analog converter on the Y axis, giving each word a unique location in the display.

*Menu.* Listing of parameter settings in software-operated logic analyzers.

*Mnemonics.* Display mode used to monitor microprocessors. Display shows the hexadecimal value plus the mnemonic code (abbreviated description of op code).

*OCT (Octal Display Mode).* The binary value of three binary channels expressed as an octal figure (0 to 7).

*Parameter Field.* That area of the display which details the parameter.

## Memory

*Compare Mode.* Comparison of new data with reference data previously stored in a second memory within the LA.

*Formattable Memory.* On some LAs it is possible to change the width and depth of the memory, e.g., halve the width and double the depth.

TABLE 19.2 ISO 7-Bit Code (646-1972)

b <sub>4</sub> b <sub>3</sub> b <sub>2</sub> b <sub>1</sub>	b <sub>7</sub> b <sub>6</sub> b <sub>5</sub>	0 0 0			0 0 1			0 1 0		
		C/R	0		1			2		
0 0 0 0	0		NUL	00	0	DLE	10	16	SP	32
0 0 0 1	1		SOH	01	1	DC1	11	17	!	33
0 0 1 0	2		STX	02	2	DC2	12	18	"	34
0 0 1 1	3		ETX	03	3	DC3	13	19	#	35
0 1 0 0	4		EOT	04	4	DC4	14	20	\$	36
0 1 0 1	5		ENQ	05	5	NAK	15	21	%	37
0 1 1 0	6		ACK	06	6	SYN	16	22	&	38
0 1 1 1	7		BEL	07	7	ETB	17	23	'	39
1 0 0 0	8		BS	08	8	CAN	18	24	(	40
1 0 0 1	9		HT	09	9	EM	19	25	)	41
1 0 1 0	10		LF	0A	10	SUB	1A	26	*	42
1 0 1 1	11		VT	0B	11	ESC	1B	27	+	43
1 1 0 0	12		FF	0C	12	FS	1C	28	,	44
1 1 0 1	13		CR	0D	13	GS	1D	29	-	45
1 1 1 0	14		SO	0E	14	RS	1E	30	.	46
1 1 1 1	15		SI	0F	15	US	1F	31	/	47

List of codes used:

NUL Null/idle  
 SOH Start of heading (CC)  
 STX Start of text (CC)  
 ETX End of text (CC)  
 EOT End of transmission (CC)  
 ENQ Enquiry (CC)  
 ACK Acknowledge (CC)  
 BEL Audible or attention signal  
 BS Backspace (FE)

HT Horizontal tabulation  
 (punch card skip) (FE)  
 LF Line feed (FE)  
 VT Vertical tabulation (FE)  
 FF Form feed (FE)  
 CR Carriage return (FE)  
 SO Shift out  
 SI Shift in  
 DLE Data link escape (CC)  
 DC1 Device control

**Memory Depth.** The number of samples that the memory can contain.

**Memory Width.** The number of channel inputs that can store data.

**Search Mode.** Philips—a very high-speed comparison mode. Others—the memory is searched for an operator-chosen word. Its location is then highlighted.

### General

**Bit.** Single unit of binary information, either 1 or 0.

**Byte.** A group of bits considered and processed together, generally 8 bits.

**CMOS.** +5-V logic; low is <1.5 V, high is >3.5 V

+10-V logic; low is <3.0 V, high is >7.0 V

+15-V logic; low is <4.5 V, high is >10.5 V

**Computer Word.** Usually 2 bytes (16 bits) but depending on system can be 8, 16, 32 bits, etc.

**ECL.** -3-V logic; low is <-1.6 V, high is >-0.75 V

**Tristate.** Used to describe a system having an intermediate logic state (high-impedance level) in addition to low and high.

**TTL.** +5-V logic; low is <0.8 V, high is >2.0 V

**Minimum Logic Swing.** Dead band around the selected threshold which the input signal must exceed to change the polarity of the comparator circuit.

### Common microprocessor terminology

**Address.** The location of a data word in the memory. The memory commences at address 0000 (hexadecimal) and continues up to FFFF (hexadecimal) with 16-bit address capabilities.

0 1 1	Hexa- decimal	Dec- imal	1 0 0	Hexa- decimal	Dec- imal	1 0 1	Hexa- decimal	Dec- imal	1 1 0	Hexa- decimal	Dec- imal	1 1 1	Hexa- decimal	Dec- imal
3			4			5			6			7		
0	30	48	@	40	64	P	50	80	'	60	96	p	70	112
1	31	49	A	41	65	Q	51	81	a	61	97	q	71	113
2	32	50	B	42	66	R	52	82	b	62	98	r	72	114
3	33	51	C	43	67	S	53	83	c	63	99	s	73	115
4	34	52	D	44	68	T	54	84	d	64	100	t	74	116
5	35	53	E	45	69	U	55	85	e	65	101	u	75	117
6	36	54	F	46	70	V	56	86	f	66	102	v	76	118
7	37	55	G	47	71	W	57	87	g	67	103	w	77	119
8	38	56	H	48	72	X	58	88	h	68	104	x	78	120
9	39	57	I	49	73	Y	59	89	i	69	105	y	79	121
:	3A	58	J	4A	74	Z	5A	90	j	6A	106	z	7A	122
:	3B	59	K	4B	75	[	5B	91	k	6B	107	{	7B	123
<	3C	60	L	4C	76	\	5C	92	l	6C	108		7C	124
	3D	61	M	4D	77	]	5D	93	m	6D	109	}	7D	125
>	3E	62	N	4E	78	ô	5E	94	n	6E	110	-	7E	126
?	3F	63	O	4F	79	—o	5F	95	o	6F	111	DEL	7F	127

DC2 Device control  
DC3 Device control  
DC4 Device control (stop)  
NAK Negative acknowledge (CC)  
SYN Synchronous idle (CC)  
ETB End of transmission block (CC)  
CAN Cancel

EM End of medium  
SUB Start of special sequence  
ESC Escape  
FS File separator (IS)  
GS Group separator (IS)  
RS Record separator (IS)  
US Unit separator (IS)  
DEL Delete

**Address Bus.** Usually 16 lines on which a binary code is placed by the microprocessor to identify a single address location.

**Assembler.** Enables programmers to write in a language one level above machine code, using mnemonic codes. The assembler software translates the mnemonics to machine code.

**Control Lines.** Logic levels on these lines signify to the rest of the system what sort of action is occurring. They are the conductors and traffic lights in the system.

**Data Bus.** Usually eight lines upon which the data being transported to or from the microprocessor are placed.

**Disassembler.** Translates instructions from machine code to mnemonics so the program can be more easily read by humans.

**Instruction Cycle.** The time required to complete one complete instruction.

**I/O.** Input-output parts of the system. Used for communication with the outside world: data are written to them and read from them in a similar way as to and from memory.

**Machine Cycle.** The time required to complete one processor action, e.g., to fetch the op code or write data to a RAM location.

**Mnemonics.** Programmers' shorthand, used when programs are written in assembly language rather than machine code.

**Multiplexed Bus Microprocessor.** The address lines are time-shared with the data lines. First, these lines contain address information that is latched into a decoder by a control line; then, they carry data information that is read with another control line.

**Program.** The program flows sequentially through the addresses unless otherwise instructed, e.g., when a certain condition arises and needs servicing by a special subprogram contained in another part of the memory.

**Program Counter.** Contains the next memory location from which to read data.

**RAM (Random-Access Memory).** Contains data and sometimes the program. Data lost if power removed.

**ROM (Read-Only Memory).** Contains the system program.

## 19.9 LOGIC PROBE

A logic probe is a small, hand-held "instrument on a chip" that is used for functional testing and troubleshooting of digital circuits. Physically, a logic probe looks like an oversized pen with a needle point at one end. The needle point is placed on various digital circuit terminals to detect the presence of logic highs or logic lows. From the opposite end of the logic probe, a strain-relieved power cord extends. This cord is connected to the power supply of the circuit under test.

A logic probe can detect the logic level at any point in a digital circuit. The display is by means of a band of light near the tip of the probe as follows: (1) a logic high is indicated by a bright light, (2) a logic low is indicated by no light, (3) an incorrect signal level produces a dim light, and (4) pulsing produces a flashing light.

A logic probe is extremely useful in tracing logic levels and pulses through IC circuitry. It can instantly show whether the mode probed is high, low, "bad level," open-circuited, or pulsing.

## 19.10 LOGIC PULSER

A logic pulser is a small, hand-held instrument on a chip. A logic pulser injects digital pulses between gates without requiring the unsoldering of components. Physically, a logic pulser looks like a logic probe. It can automatically drive low nodes high or high nodes low.

A logic pulser is a pulse generator with high-output current capability. It has sufficient current to override IC outputs in either the high or low state. It is a valuable instrument used for fault isolation or circuit design.

## 19.11 SIGNATURE ANALYSIS FOR DIGITAL TROUBLESHOOTING

Troubleshooting digital circuitry presents new challenges to the test technician. Not only have digital devices become more complex, but also the large amount of digital data generated by many of the newer designs has become difficult for the technician to interpret. The trend in the electronics industry is to replace traditional analog functions with equivalent digital circuitry. Many devices operate only at higher speeds which precludes the use of single-step testing. To detect faults of a dynamic nature, testing must be done "at speed."

A number of approaches are used to test and troubleshoot digital circuitry. Although it is not portable, the use of a computer-based system is a logical choice. The computer is capable of operating at high speeds (given the appropriate drivers and receivers to interface with the board under test) and is able to interpret the large amount of digital data often involved in testing. Using techniques to simulate the operation of the board under test, the computer can provide the user with a measure of the effectiveness of a given test. It can even be used to determine how a board will react to a hypothetical failure. Computer-based systems, however, are expensive and require programming skills and a great deal of data on the board to be tested.

A logic analyzer is a second tool used to look at digital information. Each bit of binary data is viewed on a cathode-ray tube (CRT), and in many ways a logic analyzer is equivalent to looking at waveforms on an oscilloscope. Providing the user can interpret the data, the logic analyzer provides exact information on the binary state of each bit of information for a number of parallel input lines.

## 19.12 BASIS OF SIGNATURE ANALYSIS

Signature analysis (SA) is yet another tool, and in many cases it offers a practical approach to digital troubleshooting. Arbitrarily long strings of binary data on a single node are compressed into a single, manageable signature. Signature analysis is based on the time-honored technique of signal tracing. In contrast to other digital troubleshooting methods, SA produces a signature at a node which is either good or bad. Good signatures are derived from looking at the nodes on a known good board, and all subsequent testing is done against these signatures. No analytical information is provided as to which bit in a string of binary data might be incorrect, but the advantages of SA are its simplicity and ease of operation.

Signature analysis requires that the circuitry under test be stimulated by a repetitive pattern. As a general rule, the stimulus should cause the state of every node on the board under



test to change state at least once. However, SA lacks the capability to compute how effective a given stimulus is in finding faults. In summary, the effectiveness of SA depends heavily on the quality of the stimulus to cause faults to propagate outward to one of the nodes where a measurement is made.

### 19.13 TEST ROUTINE IN ROMs

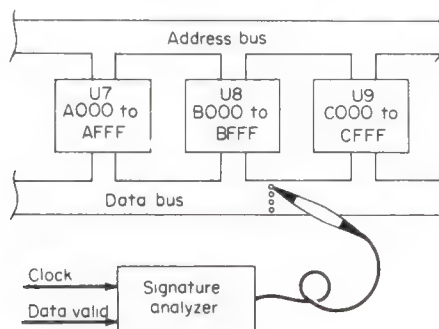
For many microprocessor-based products, manufacturers have embedded a test routine in one of the ROMs. The microprocessor and its associated memory elements can be placed into the test routine via a jumper wire. During troubleshooting, the microprocessor uses the test routine to provide a repetitive stimulus. In some cases, the microprocessor can be forced into a free-run mode by disconnecting the data bus (an NOP instruction). This forces the microprocessor into a counting routine which repeats itself. Where the manufacturer has not provided an internal stimulus, the test routine must be supplied externally by the user.

The signature is observed over a preselected window in the binary string of data at a given node. The SA is turned on at the starting point in the sequence and terminated at the ending point. The signature is derived from the digital activity which takes place during the window.

**TESTING CONTENTS OF SEVERAL ROMs** Proper selection of the window allows the user to test various portions of digital circuitry from the same node. A good example of this would be to independently test the contents of several ROMs tied to common address and data buses, as shown in Fig. 19.40. By selecting the beginning and ending addresses for a specific ROM, its contents may be viewed as signatures by probing each output node on the data bus. Once that ROM has been tested, shifting the window to include the addresses of another ROM will allow it to be tested by using the same stimulus. The stimulus need only contain the addresses of all the ROMs in the circuit to be tested (often as simple as a binary counting routine). To test U7 (in Fig. 19.40), the signature analyzer is turned on at the starting address A000 and turned off at the ending address AFFF. The contents of U8 also can be tested by prodding the same nodes on the data bus but shifting the window to a starting address of B000 and an ending address of BFFF.

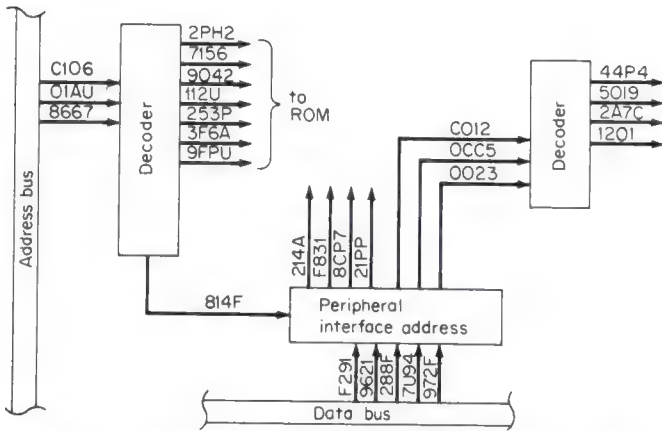
The signature analysis is synchronized to the clock rate of the board under test, and testing is accomplished dynamically at or near normal operating speed. Many SA units allow detection of unstable signatures to pinpoint timing problems. At times, unstable signatures are due to propagation delays normal to the circuit under test, and the clock must be "deskewed" by using a tapped delay line in series with the clock to the signature analyzer.

**TESTING UNKNOWN BOARDS** The entire SA technique is predicated on having good signatures from a known good board. Testing on unknown boards is accomplished by using the signatures derived from the good board. The danger in the SA approach is that small differences between boards could cause "bad" signatures on a good board. Such cases are



**Fig. 19.40** Individual ROMs can be tested by using SA.





**Fig. 19.41** Manufacturers using SA as a troubleshooting tool will typically list signatures in their operating and service manuals on a node-by-node basis.

the exception, however. As illustrated in Fig. 19.41, manufacturers list signatures on a node-by-node basis in their operating and service manuals.

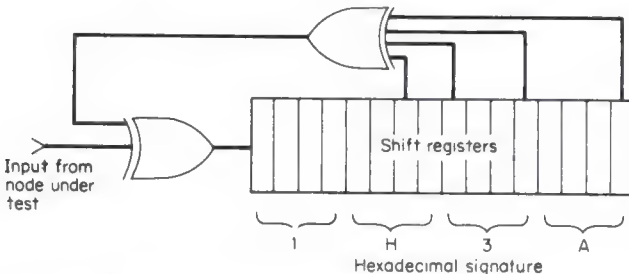
**TROUBLESHOOTING TECHNIQUE** Troubleshooting is carried out by the test technician who uses conventional back-tracing methods. If a bad signature is detected, the fault can be traced to the failing node. Because of the bus structure so common in modern digital equipment in which many devices are tied to a single set of nodes, fault isolation to a single defective component may not be possible. In other cases, it may be possible to view both signatures at the input and output nodes of a device.

#### 19.14 THE DESIGN OF A SIGNATURE ANALYZER

As illustrated in Fig. 19.42, the design of a signature analyzer is conceptually simple. The unit consists of shift registers with multiple feedback taps summed with the input data. The incoming binary data are sent through the shift registers and are modified by the feedback loops in such a way that the probability of having the same signature for two varying binary sequences is extremely low.

At the end of the window, the process of shifting data is stopped. The remainder left in the registers is the unique signature for that binary sequence. The signature itself is typically displayed as four hexadecimal characters. If 1 bit in the string is changed, the entire signature is changed.

In summary, SA is a powerful digital troubleshooting tool adaptable for field use. It does require some planning and designed-in features on the part of the manufacturer for the



**Fig. 19.42** Design of signature analyzer.

product to the tested. The technique relies on previously recorded signatures derived from a known good board. From the technician's standpoint, conventional back-tracing methods may be applied to isolate faults on digital circuitry.

#### REVIEW QUESTIONS AND ANSWERS ABOUT SIGNATURE ANALYSIS

1. What does signature analysis demand in terms of the stimulus? *Answer:* The stimulus must be repetitive and should cause the state of every node to change at least once during the test sequence.
2. Can the preselected window be changed to test other portions of the circuitry by using the same stimulus? *Answer:* Yes. Often this technique is used to test the contents of ROMs connected to a common bus.
3. Does signature analysis give the user any diagnostic information as to what bit in a binary sequence might be missing to cause an incorrect signature? *Answer:* No. Signature analysis tests on a go-no-go basis.



# Chapter 20

## Microwaves

**Richard J. Pankow**

*Senior Microwave Engineer*

*Norden Systems*

*Norwalk, Connecticut*

### 20.1 BASIC CONCEPTS

The fact that electricity does not travel instantaneously greatly affects circuits and devices in the area generally known as microwaves. The velocity of electricity, although very high ( $6 \times 10^{10}$  cm/s), results in voltages and currents being different at physically separated points. Although this phenomenon always holds, in low-frequency circuits the effects are trivial; but when the frequency of operation is sufficiently high, these effects cannot be ignored.

**WAVELENGTH** *Wavelength* is that distance traveled by electricity during the time of one complete cycle. In free space,

$$\lambda = \frac{c}{f} \quad (20.1)$$

where  $\lambda$  = wavelength  
 $c$  = speed of light  
 $f$  = frequency

**example 20.1** What is the wavelength in air of a 1-GHz ( $10^9$  Hz) wave?

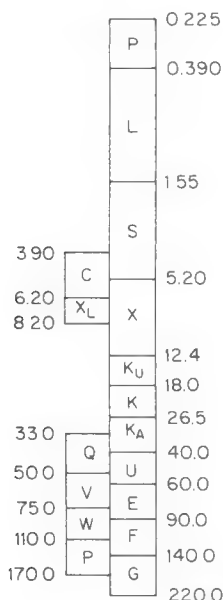
**solution** The velocity in air is essentially the same as in free space,  $2.9979 \times 10^{10}$  cm/s. So

$$\lambda = \frac{c}{f} = \frac{2.998 \times 10^{10}}{10^9} = 29.98 \text{ cm} \quad \text{Answer}$$
$$\frac{29.98}{2.54} = 11.8 \text{ in}$$

Physical devices that may occupy several inches of length must account for the delays resulting from the finite velocity of propagation.

**MICROWAVE REGION** The boundaries of the microwave region are somewhat arbitrary. It is generally agreed, however, that signals lying between 1 and 250 GHz are considered to be in the microwave region. This corresponds to wavelengths of 12 to 0.05 in (or

30.5 cm to 1.27 mm), respectively. The region is divided by convention into bands for which letters have been assigned. (See Fig. 20.1.)



**Fig. 20.1** Microwave band letter designations and frequency in gigahertz.

bel, a unit named after Alexander Graham Bell who first applied the logarithm principle to measurements he performed during his experiments on deafness. A typical microwave receiver may have an input power level of  $1 \mu\text{W}$  and an output power of  $1 \text{ W}$ , which is a gain of 1,000,000. To avoid writing large numbers, Bell conceived of using the logarithm of the number. The definition of the decibel (dB) is

$$\text{dB} = 10 \log \frac{P_2}{P_1} \quad (20.2)$$

where  $P_1$  and  $P_2$  are the powers being compared.

Some important relations which are immediately apparent are as follows:

When  $P_2$  equals  $P_1$ ,  $\text{dB} = 0$ .

When  $P_2$  is larger than  $P_1$ , the decibels are positive.

When  $P_2$  is smaller than  $P_1$ , the decibels are negative.

When  $P_2$  is 10 times (100 times)  $P_1$ ,  $\text{dB} = 10$  ( $\text{dB} = 20$ ).

When  $P_2$  is 0.1 times (0.01 times)  $P_1$ ,  $\text{dB} = -10$  ( $\text{dB} = -20$ ).

Another useful relation ensues because the logarithm of 2 is 0.3010. When the powers  $P_2$  and  $P_1$  are related by factors of 2, 4, 0.5, and 0.25, the values are +3, +6, -3, and -6 dB, respectively. In other words, when you multiply by 2, add 3 dB; when you divide by 2, subtract 3 dB.

The decibel is not an absolute unit. Rather, it indicates the relation between two powers. Decibels cannot be converted to watts except when  $P_1$  (denominator) is held constant by mutual agreement. A special name is usually assigned in such cases. The most common example is the  $\text{dB}_m$ , where the denominator is always  $1 \text{ mW}$ . A signal strength of  $1 \text{ mW}$  is  $0 \text{ dB}_m$ ,  $2 \text{ mW}$  is  $3 \text{ dB}_m$ , etc. One can use dB and  $\text{dB}_m$  for solving problems.

**example 20.2** An amplifier has an input power of  $8 \mu\text{W}$  and an output power of  $4 \text{ mW}$ . What is the gain in decibels?

**solution** Gain compares output relative to input:

### POWER AS A FUNDAMENTAL QUANTITY

Often in the measurements of microwave circuits, the concepts of voltage and current tend to lose their importance. Power is more readily measurable since it can be converted to heat. Some of the earliest microwave instruments were those designed to measure power by means of a resistive element which heated up when microwave energy impinged on it. The resulting change in resistance was sensed by a balanced Wheatstone bridge. Another early power-measuring instrument detected the temperature rise of the resistor element by means of flowing water. There were, at the time, very few instruments that could measure voltage or current directly. Furthermore, they were so large physically that they could not be used to measure a circuit without upsetting it. The evolutionary process thus begun results in many microwave components to this day being described in terms of power.

**DECIBELS** Because of the large dynamic range of power levels encountered in microwave work, it is often convenient to use the mathematical convenience of decibels. *Decibel* literally means one-tenth of a

$$\begin{aligned} \text{dB} &= 10 \log \frac{P_2}{P_1} = 10 \log \frac{4 \times 10^{-3}}{8 \times 10^{-6}} \\ &= 10 \log 500 = 10(2.7) = 27 \text{ dB} \end{aligned}$$

Answer

**example 20.3** What is the output of a 27-dB gain amplifier whose input power is 25  $\mu\text{W}$ ?

**solution**

$$\begin{aligned} \text{dB}_m &= 10 \log \frac{P_2}{1 \text{ mW}} = 10 \log \frac{25 \times 10^{-6}}{1 \times 10^{-3}} \\ &= 10 \log 0.025 = 10(-1.6) = -16 \text{ dB}_m \end{aligned}$$

So  $-16 \text{ dB}_m$  is the input power. Because output power equals input power times gain, when you use dB:

$$\begin{aligned} \text{Output power (dB}_m\text{)} &= \text{input power (dB}_m\text{)} + \text{gain (dB)} \\ P_o &= -16 + 27 = +11 \text{ dB}_m \end{aligned}$$

**example 20.4** What power in watts is equivalent to  $+11 \text{ dB}_m$ ?

**solution** Here we use the antilogarithm. Since

$$\text{dB} = 10 \log \frac{P_2}{P_1}$$

we know

$$P_2 = P_1(10^{\text{dB}/10})$$

In our case,  $P_1 = 1 \text{ mW}$  since we are dealing with  $\text{dB}_m$ . So

$$\begin{aligned} P_2 &= 1 \times 10^{-3} \times 10^{11/10} = 10^{-3} \times 10^{1.1} \\ &= 10^{-3} \times 12.6 \text{ W} \\ &= 12.6 \text{ mW} = .0126 \text{ W} \end{aligned}$$

or

Answer

When you use  $\text{dB}_w$ , it is understood that the denominator  $P_1$  is always 1 W. That is, we are always comparing the measured power to 1 W; it may be greater than, less than, or equal to 1 W.

Generally  $\text{dB}_c$  is applied when several signals coexist and one is considered the prime signal (carrier). Here we compare the relative value of the modulation, the sidebands, or any interfering signals to the desired one and express it as in decibels.

## 20.2 TRANSMISSION LINES

Any conductor or wire can be considered a transmission line. However, when we are dealing with frequencies whose wavelengths are comparable to the dimensions of the circuit, special conductors with known properties have been devised. These special conductors take several generic forms. Coaxial cable and parallel wire (such as TV lead-in) are probably the most widely known conventional types. At microwave frequencies these can be quite lossy, and other types of lines come into use. Among these are waveguides and dielectric cylinders (fiber optics).

In general, we discuss lossless lines. By this we mean that there are no dissipative elements, namely resistance. Obviously, real conductors have resistance and are lossy. However, the solutions to those problems are very complex, and as long as losses are kept low, the idealized lossless solutions give excellent results.

**CHARACTERISTIC IMPEDANCE** It is useful at this point to introduce the term *characteristic impedance*  $Z_0$ . This impedance is the ratio of the voltage to the current of a wave traveling in either direction on a transmission line. A forward and backward wave may exist simultaneously in a transmission line. At points on the line where the two waves exist simultaneously, they combine as follows:

1. Voltage waves traveling in opposite directions add.
2. Current waves traveling in opposite directions subtract.

The impedance  $Z_0$  is characteristic of the physical geometry of the transmission line. Transmission lines made of perfectly conducting wires (resistance = 0) and perfect insulat-



# 20-4      Microwaves

ing material (resistance =  $\infty$ ) are known as lossless lines. Many practical lines act almost like lossless lines.

**COAXIAL CABLE** Coaxial cable is made of two concentric conductors separated by an insulator. The velocity of propagation of an electric wavefront is given by  $V = c/\sqrt{\epsilon_r}$ , where  $\epsilon_r$  is the permittivity (dielectric constant) of the insulator. This velocity is always less than  $c$  and is independent of frequency. The characteristic impedance  $Z_0$  for a coaxial transmission line is

$$Z_0 = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{b}{a} = \frac{138}{\sqrt{\epsilon_r}} \log \frac{b}{a} \tag{20.3}$$

where  $b$  = radius of outer conductor  
 $a$  = radius of inner conductor  
 $\epsilon_r$  = relative permittivity

An abbreviated table of permittivities is shown in Table 20.1.

**example 20.5** Find the characteristic impedance of a Teflon-filled coaxial line shown below.

**solution**



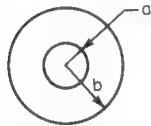
$a = 0.031$  in  
 $b = 0.250$  in

$$\begin{aligned} Z_0 &= \frac{138}{\sqrt{\epsilon_r}} \log \frac{b}{a} \\ &= \frac{138}{\sqrt{2.1}} \log \frac{0.250}{0.031} \\ &= 86.3 \, \Omega \end{aligned}$$

*Answer*

**example 20.6** For the air-filled coaxial line shown, what should the inner conductor diameter be so that  $Z_0 = 50 \, \Omega$ ?

**solution** We know that  $\epsilon_r = 1$  for air. We want to solve for  $a$ . Thus



$a = ?$   
 $b = 0.250$  in

$$\begin{aligned} Z_0 &= 138 \log \frac{b}{a} \\ \log \frac{b}{a} &= \frac{Z_0}{138} \\ \frac{b}{a} &= 10^{Z_0/138} \\ a &= \frac{b}{10^{Z_0/138}} \end{aligned}$$

**TABLE 20.1    An Abbreviated Table of Relative Permittivities**

Material	Relative dielectric constant	
	@ 1 MHz	@ 10 GHz
Free Space	1.00	1.00
Teflon	2.10	2.10
Polyethylene	2.25	2.25
Polystyrene	2.56	2.54
Nylon	3.40	3.00
Fused Quartz	3.78	3.78
Glass, Corning 7070	4.00	4.00
Bakelite	4.40	3.70
Micarta	4.50	3.20
Water	78.00	55.00

Substituting values yields

$$a = \frac{0.250}{10^{50/158}} = 0.109 \text{ in} \quad \text{Answer}$$

Coaxial cable comes in many physical forms depending on application requirements. For example, power handling, voltage breakdown, mechanical flexibility, and loss per foot must all be considered. Figure 20.2 depicts some typical forms. Table 20.2 is an abbreviated list of frequently used types of coaxial cable.

Coaxial cable is used at VHF, UHF, and the lower microwave frequencies.

**PARALLEL-WIRE TRANSMISSION LINE** As its name implies, the parallel-wire transmission line consists of two wires held at a constant spacing either by spacers placed at regular intervals or by a continuous film of insulator (TV twin lead). The characteristic impedance for an air dielectric line is

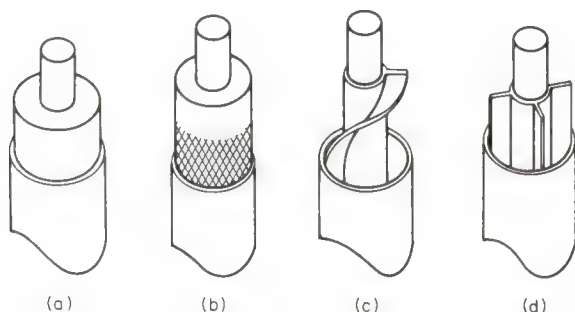
$$Z_0 = 120 \ln \frac{D}{a} = 276 \log \frac{D}{a}$$

where  $a$  is the conductor radius and  $D$  is the center-to-center spacing. The parallel-wire transmission line is used at VHF and UHF frequencies but not at microwave frequencies.

**STRIPLINE AND MICROSTRIP** Important types of transmission lines in microwave circuits besides coaxial cable are stripline and microstrip. Stripline consists of a narrow, flat conductor sandwiched between two dielectric boards whose outside surfaces are coated with conductor. Microstrip is a narrow, flat conductor on a single dielectric board whose opposite face is metallized, much like an ordinary printed-circuit board. The velocity of propagation and characteristic impedance of these types of lines are not expressible in closed form. Many handbooks contain charts for calculating these parameters from mechanical dimensions, similar to the graphs shown in Figs. 20.3 and 20.4.

**OTHER SHAPES** Occasionally, one encounters odd shapes, such as in an experimental setup, where it is desired to calculate the characteristic impedance. Figure 20.5 illustrates some of the more common ones and the characteristic impedance formulas.

**WAVEGUIDES** Waveguides come in many different shapes. The most widely used is the rectangular type, although circular, elliptical, and double-ridged shapes are available for special purposes. See Fig. 20.6. The waveguide may be made of copper, brass, or aluminum. It may have polished interior walls and sometimes may be silver-flashed. Waveguides propagate energy by bouncing the waves between the walls of the waveguide at a skew angle, like images in a house of mirrors. The concepts of voltage, current,

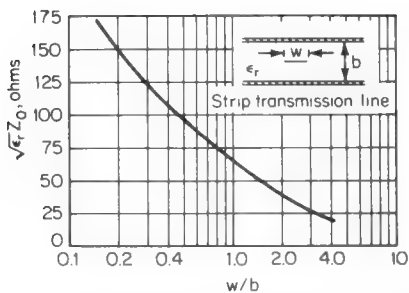


**Fig. 20.2** Typical forms of coaxial cable. (a) Semirigid, thin-walled copper outer conductor with solid dielectric insulator. (b) Flexible braided outer conductor with vinyl jacket. (c) Helical spacer for inner conductor. (d) Radial spacer for inner conductor.

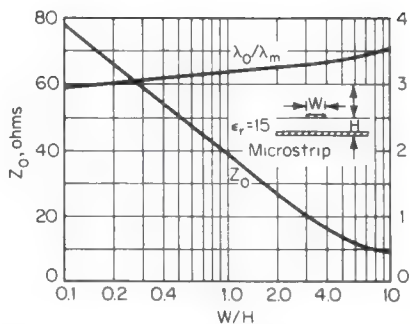


62/U	9863, 9864, 9865	84.0	93.0	0.90	1.90	2.80	3.70	5.80	8.50	—
62A/U	9062A	84.0	93.0	0.90	1.90	2.80	3.70	5.80	8.50	—
63B/U	9063B	84.0	125.0	0.50	1.10	1.50	2.30	3.40	5.70	—
71B/U	9071B	84.0	93.0	0.90	1.90	2.80	3.70	5.80	8.50	18.60
174/U	9174	66.0	50.0	3.80	6.50	8.90	12.00	17.50	31.00	—
178B/U	9178B	70.0	50.0	5.30	10.00	13.30	20.00	27.50	45.00	—
179B/U	9179B	70.0	75.0	5.00	7.90	9.80	12.70	15.80	25.00	—
180B/U	9180B	70.0	93.0	3.10	4.20	5.10	7.30	10.40	16.50	—
187A/U	9187A	70.0	75.0	5.00	7.90	9.80	12.70	15.80	25.00	—
188A/U	9188A	70.0	50.0	3.80	7.90	11.50	15.00	20.00	30.00	58.00
195A/U	9195A	70.0	95.0	3.10	4.20	5.10	7.30	10.40	16.50	—
196A/U	9196A	70.0	50.0	5.30	10.00	13.30	20.00	27.50	45.00	—
212/U	9212	66.0	50.0	0.65	1.60	2.40	3.60	5.20	8.80	16.70
213/U	9213	66.0	50.0	0.66	1.50	2.20	3.20	4.60	9.00	19.00
214/U	9214	66.0	50.0	0.66	1.50	2.20	3.20	4.60	9.00	19.00
215/U	9215	66.0	50.0	0.66	1.50	2.20	3.20	4.60	9.00	19.00
217/U	9217	66.0	50.0	0.41	1.00	1.40	2.10	3.10	5.80	13.00
218/U	9218	66.0	50.0	0.23	0.56	0.81	1.20	1.90	3.80	9.00
219/U	9219	66.0	50.0	0.23	0.56	0.81	1.20	1.90	3.80	9.00
223/U	9223	66.0	50.0	1.35	3.00	4.30	6.00	8.80	16.50	—
AlphaMini	9450	70.0	50.0	6.50	13.00	16.00	25.50	33.00	—	—
AlphaMini	9806	78.0	93.0	3.10	4.20	5.10	7.30	10.40	—	—
AlphaMini	9475	80.0	75.0	5.00	7.90	9.80	12.70	15.80	—	—
AlphaMini	9500	80.0	100.0	4.50	7.40	9.60	12.50	15.70	—	—

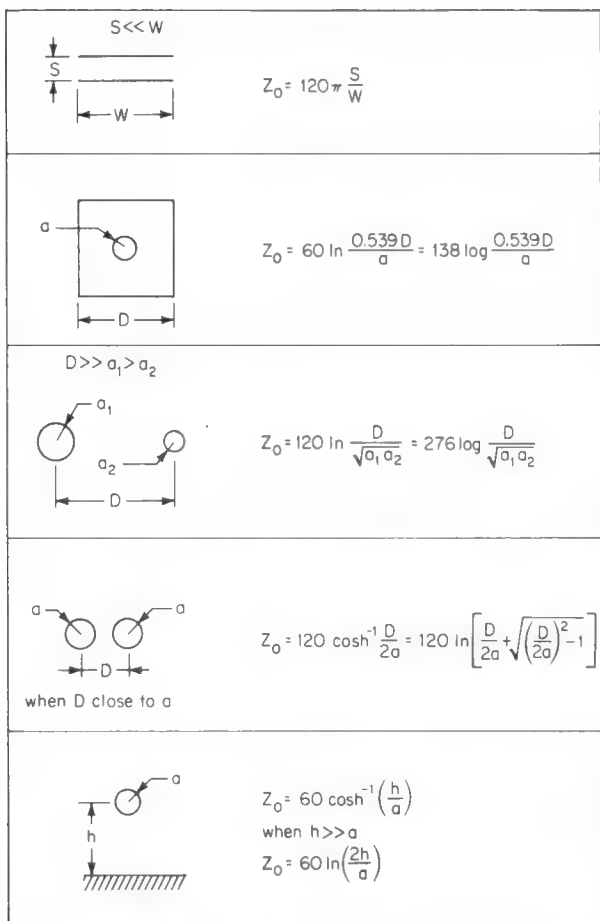
Note: + = precision video cable,  
+ = twinaxial construction.  
X = plenum cable.



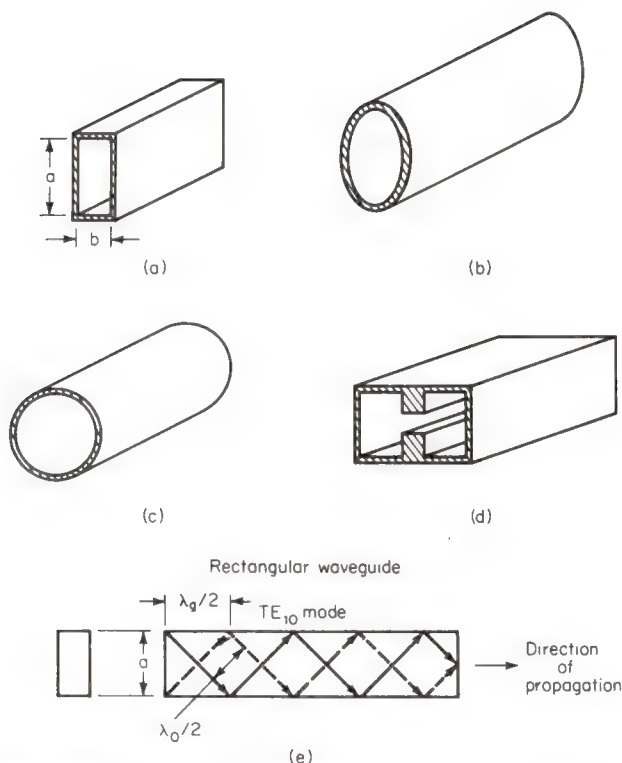
**Fig. 20.3** Characteristic impedance of strip transmission line for the special case of a very thin center strip. (After S. B. Cohn, "Problems in Strip Transmission Lines," *IRE Transactions on Microwave Theory and Techniques*, vol. 3, no. 2, March 1955.)



**Fig. 20.4** Characteristic impedance and  $\lambda_0/\lambda_m$  ratio for microstrip transmission line. (From the 1969 *Microwave Engineers' Handbook and Buyers' Guide*, as presented by Burke, Gelnovatch, and Chase, based on Wheeler's work.)



**Fig. 20.5** Unusual shapes for transmission lines and their characteristic impedance formulas.



**Fig. 20.6** Waveguides: (a) Rectangular, (b) elliptical, (c) circular, (d) double-ridged, (e) method of propagation.

characteristic impedance, and velocity of propagation are far less useful in understanding the waveguide than in the previously discussed transmission lines. Waveguides are less lossy at a given frequency than an equal length of coaxial cable. Waveguide is more restricted in its range of frequency usefulness than is coaxial cable. For example, RG-8B/U, a popular coaxial cable, has a loss of 17 dB per 100 ft at 3 GHz whereas WR-284 waveguide loss per 100 ft is about 0.8 dB at the same frequency, a reduction of 20 to 1. However, the recommended frequency range for the waveguide is 2.60 to 3.95 GHz, as shown in Table 20.3 while RG-8B/U cable is useful from direct current to 10 GHz. The reason for the narrow-band nature of the waveguide lies in the method of energy propagation. When the free-space wavelength is too large, i.e., the frequency is too low, the wave does not “fit” inside the guide. As an analogy, compare AM and FM reception when you are driving through a short tunnel:

Wavelength of AM	980 ft
Wavelength of FM	9.8 ft
Diameter of tunnel	37 ft

Specifically, a wave will not propagate in a rectangular waveguide if the frequency is less than the cutoff frequency

$$f_{co} = \frac{c}{2a}$$

where  $c$  is speed of light in air (air-filled waveguide) and  $a$  is the larger inside dimension of the waveguide. At the high-frequency end, the wavelength becomes sufficiently small that more than one wavelength can fit into the waveguide. This is known as higher-order modes and is undesirable.



**TABLE 20.3 Reference Table of Rigid Rectangular Waveguide Data and Fittings**  
(Courtesy Microwave Development Laboratories, Inc., Natick, Mass.)

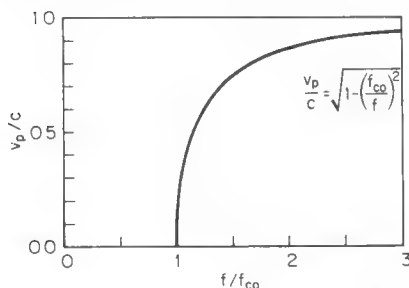
EIA WG designation WR-( )	Recommended operating ranges for TE <sub>10</sub> mode		Cutoff for TE <sub>10</sub> mode		Range $\frac{2\lambda}{\lambda_c}$ in $\frac{a}{\lambda_c}$	Range $\frac{\lambda_g}{\lambda}$ in $\frac{a}{\lambda}$	Theoretical CW power rating lowest to highest frequency (MW)	Theoretical attenuation lowest to highest frequency (db/100ft)
	Frequency (GHz)	Wavelength (cm)	Frequency (GHz)	Wavelength (cm)				
2300	0.32-0.49	93.68-61.18	0.256	116.84	1.60-1.05	1.68-1.17	153 0-212.0	051-031
2100	0.35-0.53	85.65-56.56	0.281	106.68	1.62-1.06	1.68-1.18	120.0-173.0	054-034
1800	0.41-0.625	73.11-47.96	0.328	91.44	1.60-1.05	1.67-1.18	93.4-131.9	056-038
1500	0.49-0.75	61.18-39.97	0.393	76.20	1.61-1.05	1.62-1.17	67.6-93.3	069-050
1150	0.64-0.96	46.84-31.23	0.513	58.42	1.60-1.07	1.82-1.18	35.0-53.8	128-075
975	0.75-1.12	39.95-26.76	0.605	49.53	1.61-1.08	1.70-1.19	27.0-38.5	137-095
770	0.96-1.45	31.23-20.67	0.766	39.12	1.60-1.06	1.66-1.18	17.2-24.1	201-136
650	1.12-1.70	26.76-17.63	0.908	33.02	1.62-1.07	1.70-1.18	11.9-17.2	317-212
								269-178
510	1.45-2.20	20.67-13.62	1.157	25.91	1.60-1.05	1.67-1.18	7.5-10.7	
430	1.70-2.60	17.63-11.53	1.372	21.84	1.61-1.06	1.70-1.18	5.2-7.5	588-385
								501-330
340	2.20-3.30	13.63-9.08	1.736	17.27	1.58-1.05	1.78-1.22	3.1-4.5	877-572
								751-492
284	2.60-3.95	11.53-7.59	2.078	14.43	1.60-1.05	1.67-1.17	2.2-3.2	1 102-752
								940-641
229	3.30-4.90	9.08-6.12	2.577	11.63	1.56-1.05	1.62-1.17	1.6-2.2	
187	3.95-5.85	7.59-5.12	3.152	9.510	1.60-1.08	1.67-1.19	1.4-2.0	2 08-1.44
								1.77-1.12
159	4.90-7.05	6.12-4.25	3.711	8.078	1.51-1.05	1.52-1.19	0.79-1.0	
137	5.85-8.20	5.12-3.66	4.301	6.970	1.47-1.05	1.48-1.17	0.56-0.71	2 87-2.30
								2.45-1.94
112	7.05-10.00	4.25-2.99	5.259	5.700	1.49-1.05	1.51-1.17	0.35-0.46	4.12-3.21
								3.50-2.74
90	8.20-12.40	3.66-2.42	6.557	4.572	1.60-1.06	1.68-1.18	0.20-0.29	6.45-4.45
								5.49-3.83
75	10.00-15.00	2.99-2.00	7.868	3.810	1.57-1.05	1.64-1.17	0.17-0.23	
62	12.4-18.00	2.42-1.66	9.486	3.160	1.53-1.05	1.55-1.18	0.12-0.16	9.51-8.31
								6.14-5.36
51	15.00-22.00	2.00-1.36	11.574	2.590	1.54-1.05	1.58-1.18	0.080-0.107	
42	18.00-26.50	1.66-1.13	14.047	2.134	1.56-1.06	1.60-1.18	0.043-0.058	20.7-14.8
								17.6-12.6
								13.3-9.5
34	22.00-33.00	1.36-0.91	17.328	1.730	1.57-1.05	1.62-1.18	0.034-0.048	
28	26.50-40.00	1.13-0.75	21.081	1.422	1.59-1.05	1.65-1.17	0.022-0.031	21.9-15.0
22	33.00-50.00	0.91-0.60	26.342	1.138	1.60-1.05	1.67-1.17	0.014-0.020	31.0-20.9
19	40.00-60.00	0.75-0.50	31.357	0.956	1.57-1.05	1.63-1.16	0.011-0.015	
15	50.00-75.00	0.60-0.40	39.863	0.752	1.60-1.06	1.67-1.17	0.0063-0.0090	52.9-39.1
12	60.00-90.00	0.50-0.33	48.350	0.620	1.61-1.06	1.68-1.18	0.0042-0.0060	93.3-52.2
10	75.00-110.00	0.40-0.27	59.010	0.508	1.57-1.05	1.61-1.18	0.0030-0.0041	
8	90.00-140.00	0.333-0.214	73.840	406	1.64-1.05	1.75-1.17	0.0018-0.0026	152-99
7	110.00-170.00	0.272-0.176	90.840	330	1.64-1.06	1.77-1.18	0.0012-0.0017	163-137
5	140.00-220.00	0.214-0.136	115.750	259	1.65-1.05	1.78-1.17	0.00071-0.00107	308-193
4	170.00-260.00	0.176-0.115	137.520	218	1.61-1.05	1.69-1.17	0.00052-0.00075	384-254
3	220.00-325.00	0.136-0.092	173.280	173	1.57-1.06	1.62-1.18	0.00035-0.00047	512-348

Material alloy	JAN flange designation			EIA WG designations WR- ( )	Dimensions (in)				Wall thickness nominal
	JAN WG designations RG- ( )/U	Choke UG( )/U	Cover UG( )/U		Inside	Tolerance	Outside	Tolerance	
Alum.	290			2300	23.000-11.500	± .020	23.250-11.750	± .020	0.125
Alum.	291			2100	21.000-10.500	± .020	21.250-10.750	± .020	0.125
Alum.	201			1800	18.000-9.000	± .020	18.250-9.250	± .020	0.125
Alum.	202			1500	15.000-7.500	± .015	15.250-7.750	± 0.15	0.125
Alum.	203			1150	11.500-5.750	± .015	11.750-6.000	± .015	0.125
Alum.	204			975	9.750-4.875	± .010	10.000-5.125	± .010	0.125
Alum.	205			770	7.700-3.850	± .005	7.950-4.100	± .005	0.125
Brass	89								
Alum.	103		417A	650	6.500-3.250	± .005	6.660-3.410	± .005	0.080
Brass	337		418A	510	5.100-2.550	± .005	5.260-2.710	± .005	0.080
Brass	104		435A	430	4.300-2.150	± .005	4.460-2.310	± .005	0.080
Alum.	105		437A						
Brass	112		553	340	3.400-1.700	± .005	3.560-1.860	± .005	0.080
Alum.	113		554						
Brass	48	54A	53	284	2.840-1.340	± .005	3.000-1.500	± .005	0.080
Alum.	75	585	584	229	2.290-1.145	± .005	2.418-1.273	± .005	0.064
Brass	340			187	1.872-0.872	± .005	2.000-1.000	± .005	0.064
Brass	49	148B	149A	159	1.590-0.795	± .004	1.718-0.923	± .004	0.064
Alum.	95	406A	407	137	1.372-0.622	± .004	1.500-0.750	± .004	0.064
Brass	343								
Brass	50	343A	344	112	1.122-0.497	± .004	1.250-0.625	± .004	0.064
Alum.	106	440A	441	90	0.900-0.400	± .003	1.000-0.500	± .003	0.050
Brass	51	52A	51	75	0.750-0.375	± .003	0.850-0.475	± .003	0.050
Alum.	68	137A	138	62	0.622-0.311	± .0025	0.702-0.391	± .003	0.040
Brass	52	40A	39						
Alum.	67	136A	135						
Brass	346								
Brass	91	541	419	51	0.510-0.255	± .0025	0.590-0.335	± .003	0.040
Alum.	349	—	—	42	0.420-0.170	± .0020	0.500-0.250	± .003	0.040
Silver	107	—	—						
Brass	352								
Brass	53	596	595						
Alum.	121	59M	597						
Silver	66	—	—						
Brass	354	600	599	34	0.340-0.170	± .0020	0.420-0.250	± .003	0.040
Alum.	355	—	—	28	0.280-0.140	± .0015	0.360-0.220	± .002	0.040
Silver	96	—	—						
Brass	271		383	22	0.224-0.112	± .0010	0.304-0.192	± .002	0.040
Silver	97		—	19	0.188-0.094	± .0010	0.268-0.174	± .002	0.040
Brass	358			15	0.148-0.074	± .0010	0.228-0.154	± .002	0.040
Brass	273		385						
Silver	98		—						
Brass	274		37	12	0.122-0.061	± .0005	0.202-0.141	± .002	0.040
Silver	99		—						
Brass	359			10	0.100-0.050	± .0005	0.180-0.130	± .002	0.040
Silver	138	—	—	8	0.080-0.040	± 0.0003	0.156 DIA	± .001	—
Silver	136	—	—	7	0.065-0.0325	± 0.00025	0.156 DIA	± .001	—
Silver	135	—	—	5	0.051-0.0255	± 0.00025	0.156 DIA	± .001	—
Silver	137	—	—	4	0.043-0.0215	± 0.00020	0.156 DIA	± .001	—
Silver	139	—	—	3	0.034-0.0170	± 0.00020	0.156 DIA	± .001	—

	General
Cutoff frequency (air-filled)	$f_{co} = \frac{c}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$ <p>For TE modes  <math>m = 0, 1, 2, \dots</math>  <math>n = 0, 1, 2, \dots</math>  but <math>m \neq n</math> if either = 0</p> <p>For TM modes  <math>m = 1, 2, 3, \dots</math>  <math>n = 1, 2, 3, \dots</math></p>
	TE <sub>10</sub> mode
Cutoff frequency	$f_{co} = \frac{c}{2a}$
Velocity of axial propagation (group velocity)	$v_p = c \sqrt{1 - \left(\frac{f_{co}}{f}\right)^2}$
Phase velocity	$v_\phi = \frac{c}{\sqrt{1 - (f_{co}/f)^2}}$
Guide wavelength	$\lambda_g = \frac{\lambda_0}{\sqrt{1 - (f_{co}/f)^2}}$

**Fig. 20.7** Key rectangular waveguide equations.

The velocity of propagation through a waveguide varies with frequency while in coaxial cable it does not. Again, this is due to the nature of the propagation method. Just above the cutoff frequency, energy begins to progress down the guide, and the velocity approaches the speed of light as frequency increases. As previously stated, waveguides are generally not used above the next higher-order mode frequency. See Figs. 20.7 and 20.8.

**Fig. 20.8** Velocity of propagation in waveguide TE<sub>10</sub> mode.

**example 20.7** Waveguide type WR-62 is propagating a 16.5-GHz signal in the TE<sub>10</sub> mode. To make slotted-line measurements, we need to find the guide wavelength.

**solution** The free-space wavelength is

$$\lambda_0 = \frac{c}{f} = \frac{11.8}{16.5} = 0.715 \text{ in}$$

$$c = 3 \times 10^{10} \text{ cm/s} = 11.8 \text{ in/ns}$$

or

$$c = 11.8 \text{ in/ns}$$

and the cutoff frequency is

$$f_{co} = \frac{c}{2a} = \frac{11.8 \times 10^9}{2(0.622)} = 9.486 \text{ GHz}$$

So

$$\lambda_g = \frac{\lambda_0}{\sqrt{1 - (f_{co}/f)^2}} = \frac{0.715}{\sqrt{1 - (9.486/16.5)^2}} = 0.874 \text{ in}$$

What is the next higher-order mode frequency?

$$f_{co} = \frac{c}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$$

$$\text{For TE}_{20}: f_{co} = \frac{11.8 \times 10^9}{2} \sqrt{\left(\frac{2}{0.622}\right)^2 + 0} = 18.9 \text{ GHz}$$

$$\text{For TE}_{11}: f_{co} = \frac{11.8 \times 10^9}{2} \sqrt{\left(\frac{1}{0.622}\right)^2 + \left(\frac{1}{0.311}\right)^2} = 21.21 \text{ GHz}$$

For a piece of this waveguide 100 ft long, what delay is caused, compared to free space?

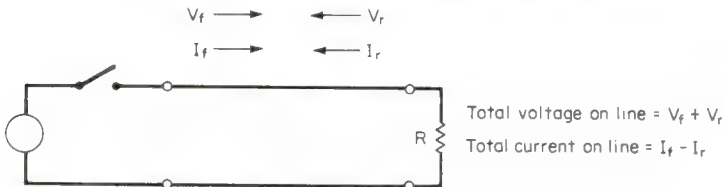
$$v_p = c \sqrt{1 - \left(\frac{f_{co}}{f}\right)^2} = 11.8 \times 10^9 \sqrt{1 - \left(\frac{9.486}{16.5}\right)^2} = 9.655 \times 10^9 \text{ in/s}$$

$$t = \frac{d}{v_p} = \frac{100 \times 12}{9.655 \times 10^9} = 124.3 \text{ ns} \quad \text{through waveguide}$$

$$t = \frac{d}{c} = \frac{100 \times 12}{11.8 \times 10^9} = 101.7 \text{ ns} \quad \text{free space}$$

**DIELECTRIC TRANSMISSION LINES** Dielectric transmission line (fiber optics, inverse guide) is similar to waveguide in its method of transporting energy; the difference lies in the surface used to bounce the waves within the guide. Although in waveguide conducting metals are used as boundaries, here total internal reflection (as in a prism) is used to confine energy. These transmission lines most often are reserved for frequencies above 100 GHz and quasi-optical energy. An exception is the commonly used dielectric rod or polyrod antenna, where because of the lower frequency the internal reflections are not total and energy is radiated into space.

**REFLECTION COEFFICIENT** The mechanism by which a load at the end of transmission line communicates its presence to the source is the reflection coefficient. In Fig. 20.9, when the switch closes at time zero, a voltage and current wave start down the transmission line



$$\frac{V_f}{I_f} = Z_0 \quad \frac{V_r}{I_r} = Z_0 \quad \frac{V}{I} = R$$

$$\frac{V}{I} = R = \frac{V_f + V_r}{I_f - I_r} = \frac{V_f + V_r}{\frac{V_f + V_r}{Z_0} - \frac{V_r - V_f}{Z_0}} = Z_0 \frac{V_f + V_r}{V_f - V_r} = Z_0 \frac{1 + V_r/V_f}{1 - V_r/V_f}$$

$$\frac{V_r}{V_f} = \frac{R - Z_0}{R + Z_0} \equiv \Gamma_V$$

**Fig. 20.9** Illustration of the reflection coefficient.

toward the load. As stated previously, they are related by the characteristic impedance and have no idea what load they will encounter. If, as in the figure, the line is terminated with a resistor, the voltage and current waves must satisfy two conditions simultaneously at the resistor: first, Ohm's law, which says that the voltage divided by the current must equal the value  $R$ ; second, the quotient of a voltage-to-current wave must equal  $Z_0$ . For both conditions to be satisfied simultaneously, one might hypothesize the generation of a backward-going voltage and current wave. While this may seem like a giant mental leap, it satisfies the mathematical criteria of solution in a simple way and does not contradict any physical observations. The backward-traveling wave, sometimes called the *reflected* wave, is related to the forward or *incident* wave by the factor  $\Gamma_v$  as derived in Fig. 20.9. Almost always one speaks of the voltage reflection coefficient. It is easy to show that the relation of the forward and backward current waves is the negative of the forward and backward voltage waves:  $\Gamma_i = -\Gamma_v$ .

**example 20.8** Find the voltage reflection coefficient  $\Gamma_v$  for a  $50\ \Omega = Z_0$  line terminated with (a) a  $100\ \Omega$  resistor load, (b) a  $25\ \Omega$  resistor load.

**solution** For part (a) of the example,

$$\Gamma_v = \frac{R - Z_0}{R + Z_0} = \frac{100 - 50}{100 + 50} = \frac{1}{3} \quad \text{Answer}$$

For part (b) of the example,

$$\Gamma_v = \frac{R - Z_0}{R + Z_0} = \frac{25 - 50}{25 + 50} = \frac{-25}{75} = -\frac{1}{3} \quad \text{Answer}$$

It is easy to show that for any passive value of resistance (a positive value) the reflection coefficient ranges from  $-1$  at  $R = 0$  to  $+1$  at  $R = \infty$ .

When the excitation is a sinusoid, the principle of the reflection coefficient still applies. If the load has reactive components, then  $\Gamma_v$  takes on complex values.

**example 20.9** Find  $\tilde{\Gamma}_v$  for a load of  $\tilde{Z}_L = 50 + j50$  on a line  $50 = Z_0$ .

**solution**

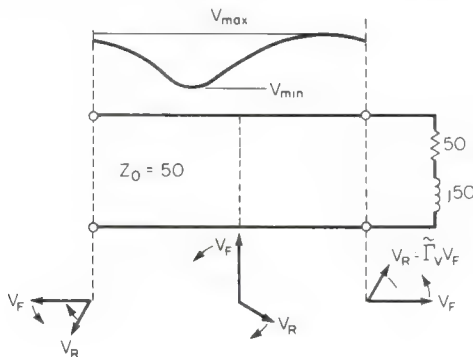
$$\begin{aligned} \tilde{\Gamma}_v &= \frac{\tilde{Z}_L - Z_0}{\tilde{Z}_L + Z_0} = \frac{50 + j50 - 50}{50 + j50 + 50} = \frac{j50}{100 + j50} = \frac{50/90^\circ}{111.8/26.6^\circ} \\ &= 0.45/63.4^\circ = \tilde{\Gamma}_v \\ &= 0.20 + j0.40 = \tilde{\Gamma}_v \end{aligned}$$

Answer

Answer

**VOLTAGE STANDING WAVE RATIO** When forward and backward sinusoidal traveling waves exist simultaneously, they create an interference pattern, because at different locations on the line the sine waves will be in phase, out of phase, or somewhere in between. The ratio of the largest magnitude voltage to the smallest magnitude voltage is called the VSWR. Note that VSWR relates magnitudes only and so can never be a complex number like the reflection coefficient. One can, however, establish a relationship between the magnitude of  $\Gamma_v$  and VSWR, as shown in Example 20.10.

**example 20.10** What is the VSWR on the  $50\text{-}\Omega$  line terminated as shown?



**solution** From Example 20.9,

$$\tilde{\Gamma}_v = 0.20 + j0.40 \quad \text{or} \quad \tilde{\Gamma}_v = 0.45/\underline{63.4^\circ}$$

$$\text{Since} \quad V_{\max} = |V_F| + |V_R| = |V_F| + |\tilde{\Gamma}_v| |V_F| = |V_F| (1 + |\tilde{\Gamma}_v|)$$

$$\text{similarly,} \quad V_{\min} = |V_F| (1 - |\tilde{\Gamma}_v|)$$

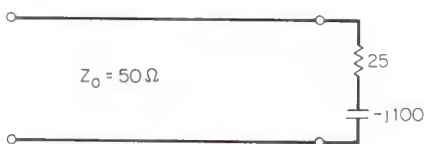
$$\text{So} \quad \text{VSWR} = \frac{V_{\max}}{V_{\min}} = \frac{1 + |\tilde{\Gamma}_v|}{1 - |\tilde{\Gamma}_v|}$$

$$\text{Now } |\tilde{\Gamma}_v| = 0.45, \text{ so} \quad \text{VSWR} = \frac{1 + 0.45}{1 - 0.45} = 2.64 \quad \text{Answer}$$

Historically, measurements of VSWR were the first to be done because the apparatus required is quite simple. By measuring the VSWR and the location of the nulls of the interference pattern, one could determine algebraically the nature and value of an unknown reflection coefficient.

**SMITH CHART** Phillip H. Smith devised a transmission line chart in 1939 which bears his name. It was an improvement on the charts that existed previously for microwave work. One of its primary uses is to quickly translate VSWR measurements to values of reflection coefficient and load impedance. The chart also finds great application in the area of filter design. Remember when you use the chart that all impedances are expressed in a normalized fashion, that is, the actual impedance is divided by  $Z_0$ . Position on a transmission line is determined by rotating a point about the center of the chart; once around corresponds to a one-half wavelength distance. Several examples best illustrate its use.

**example 20.11** See Fig. 20.10. An impedance of  $Z_L = 25 - j100$  terminates a  $50\text{-}\Omega$  transmission line. What is the reflection coefficient  $\Gamma_v$ ? What is the VSWR?



**solution**

Step 1. Normalize to the line impedance:

$$Z_L' = \frac{Z_L}{Z_0} = 0.5 - j2$$

Step 2. Plot the point  $P$  on the chart and draw circle (A) through the point centered on the chart.

Step 3. Where circle A crosses the diameter, project down and read  $|\Gamma_v|$  and VSWR:

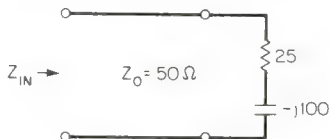
$$|\Gamma_v| = 0.82$$

$$\text{VSWR} = 10$$

Step 4. To determine the angle of  $\Gamma_v$ , draw radius  $B$  through point  $P$  and read the angle:

$$\angle \Gamma_v = -51^\circ$$

**example 20.12** See Fig. 20.11. By putting an element in shunt with the transmission line at some point, we may cause the input impedance of the line to appear real and equal to  $50\text{ }\Omega$ . Find the value and location of the shunt impedance required, which is closest to the load end of the line.





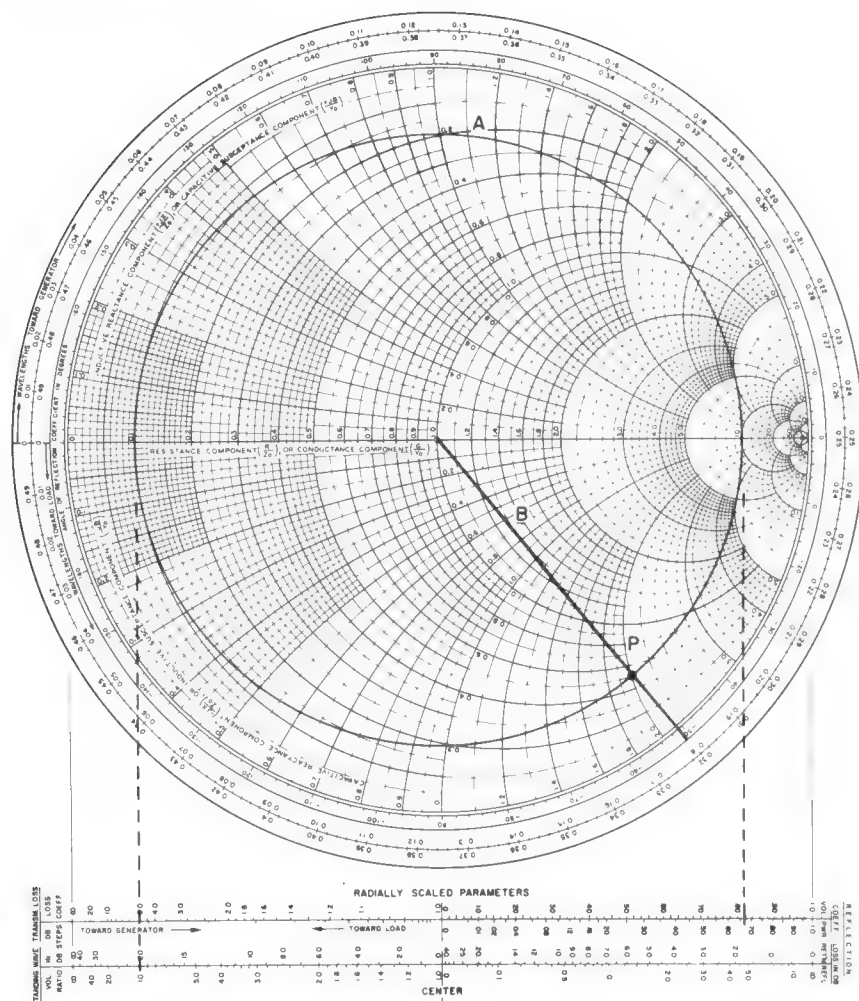


Fig. 20.10 Smith chart for Example 20.11. (Courtesy Analog Instruments, New Providence, N.J.)

### solution

Step 1. Since we are going to use a shunt element, convert to admittance. Plot the admittance point  $Q$  and draw a circle through it:

$$Z_L = 25 - j100 \quad \frac{Z_L}{Z_0} = 0.5 - j2 = Z'_L \quad Y'_L = \frac{1}{Z'_L} = \frac{1}{0.5 - j2} = 0.12 + j0.47$$

(Note:  $Q$  is diametrically opposed to  $Z'_L = 0.5 - j2$ .)

Step 2. The first point toward generator (CW) where the circle intersects the circle  $G = 1$  is point  $W$ . At this point on the line, the input admittance is  $1 + j2.9$ . The distance along the line away from the load is read off the chart as  $0.202 - 0.07 = 0.132$  wavelength.

Step 3. By placing a shunt susceptance of  $-j2.9$  at this point, the net input admittance is

$$Y' = 1 + j2.9 - j2.9 = 1 \quad \therefore Z' = 1 \quad \text{or} \quad Z = 50 \Omega$$

Step 4. A susceptance of  $-j2.9$  equals a reactance of  $1/(-j2.9) = j.345$ , which is inductive reactance.

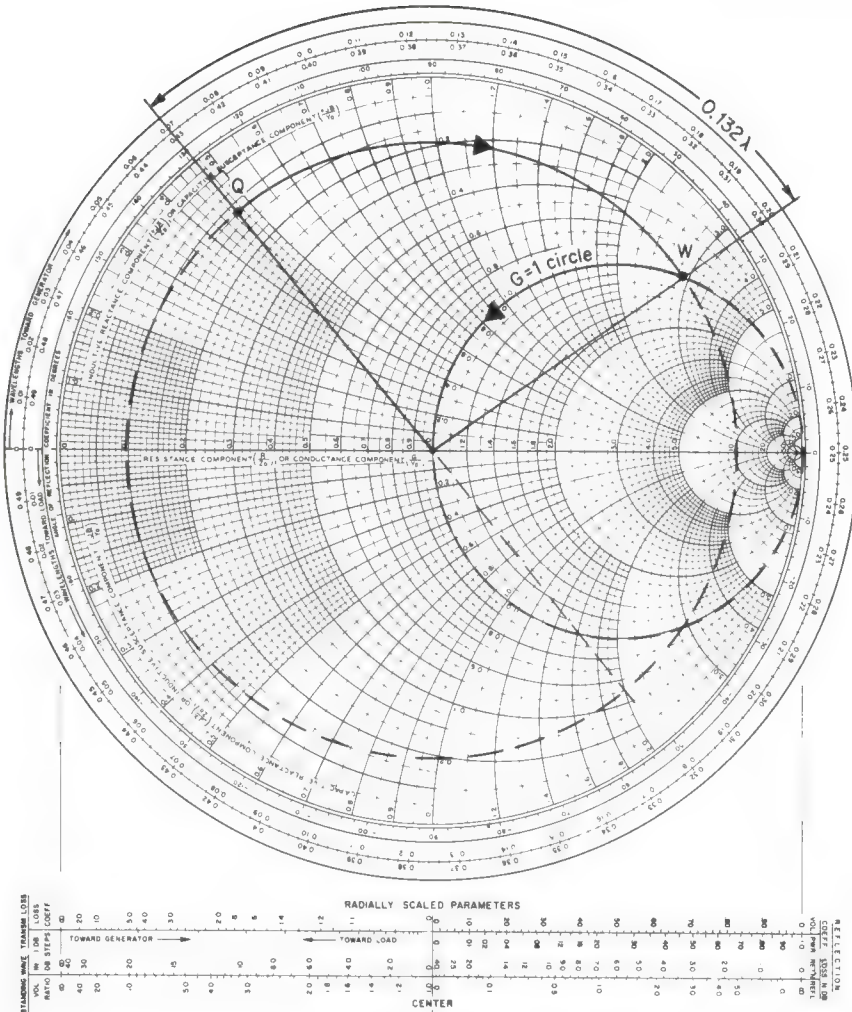
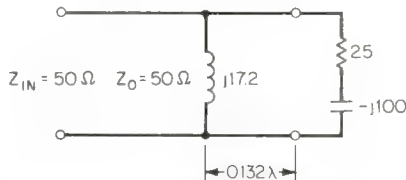


Fig. 20.11 Smith chart for Example 20.12. (Courtesy Analog Instruments, New Providence, N.J.)

Step 5. To unnormalize, multiply by  $Z_0$ :

$$X = j17.2$$



**example 20.13** See Fig. 20.12. Slotted-line measurements on a WR-90 waveguide slotted line at 10 GHz show a minimum voltage at 4.64 cm from the reference plane. A VSWR of 3.0 was measured also. What is the impedance at the reference plane?

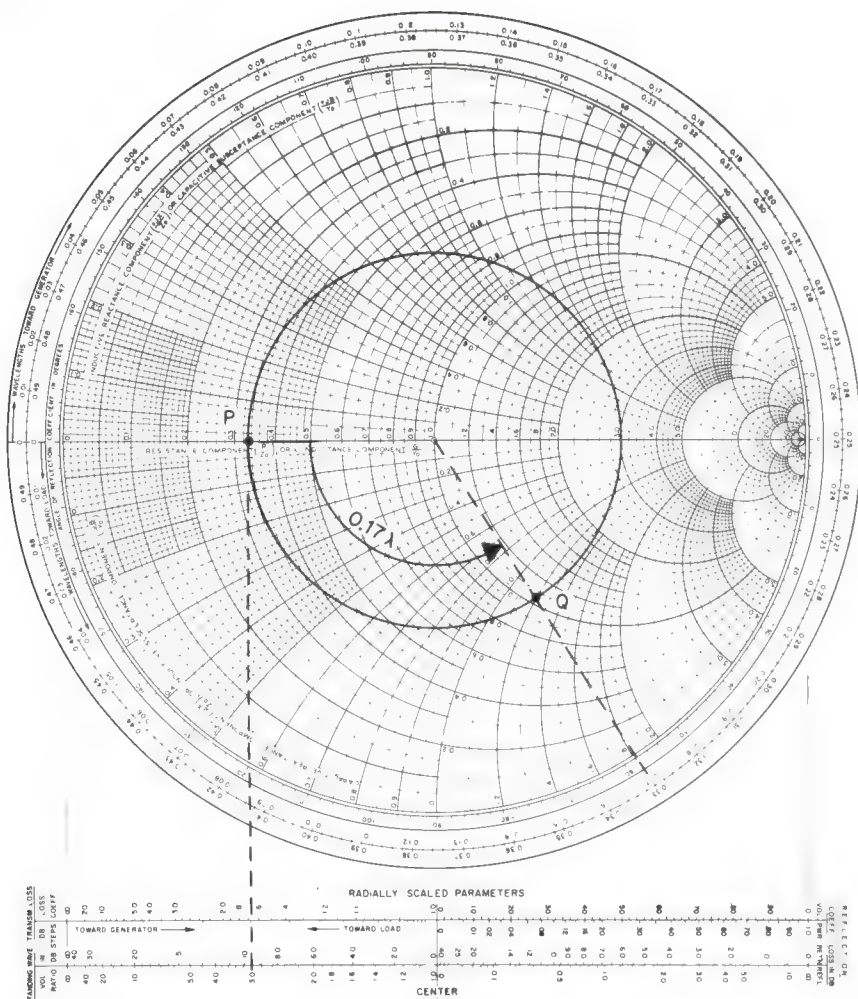


Fig. 20.12 Smith chart for Example 20.13. (Courtesy Analog Instruments, New Providence, N.J.)

### solution

Step 1. Calculate the guide wavelength  $\lambda_g$ :

$$\lambda_g = \frac{\lambda_0}{\sqrt{1 - (f_{co}/f)^2}}$$

From Table 17.3  $f_{co} = 6.557$  GHz, and

$$\lambda_0 = \frac{3 \times 10^{10}}{10 \times 10^9} = 3 \text{ cm}$$

so

$$\lambda_g = \frac{3}{\sqrt{1 - (6.557/10)^2}} = 3.97 \text{ cm}$$

Step 2. Locate the VSWR on the scale of the Smith chart, and draw the circle of this radius.

Step 3. The distance of 4.64 cm corresponds to  $4.64/3.97 = 1.17$  wavelengths. Starting at the minimum voltage point  $P$ , move toward load (CCW) by 1.17 wavelengths. That is twice around the chart plus 0.17 wavelength.

Step 4. Read off the normalized impedance at  $Q$ :

$$Z_L = 1.05 - j1.2$$

**POWER TRANSFER** Most of us are already familiar with the principle of matching load impedance to source impedance for maximum power delivered to the load. At microwave frequencies when a transmission line is placed between source and load, that principle is easily extended. For maximum power to be delivered to the load, the source impedance, transmission line impedance, and load impedance should all be equal. Since the reflection coefficient for a resistor of value  $Z_0$  is 0, all power is absorbed by it and none is reflected. At the input end to such a line, an impedance  $Z_0 = R$  is presented to the source.

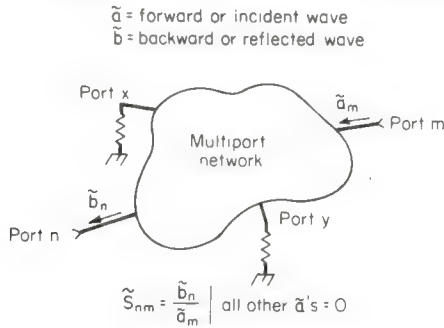


Fig. 20.13 A generalized S parameter concept.

**S PARAMETERS** With the advent of more sophisticated measuring equipment for microwave work, the S parameter method of characterizing circuits has gained popularity. Microwave circuits are categorized by the number of ports they have. A port might be a coaxial connector or a waveguide opening. To characterize a circuit means to quantify the effects of all other ports on a particular one until all have been quantified. A one-port circuit needs only one such parameter to describe it completely, a two-port circuit requires four, a three-port nine, etc. The four S parameters of a two-port circuit completely describe in both magnitude and phase how that circuit will react to any combination of input and output signals. The simplest example of an S parameter is the case of a one-port circuit where the complex reflection coefficient  $\Gamma_v$  is the only S parameter required.

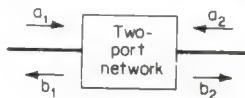
The S parameters are the magnitude and phase relations between voltage waves. They are of two types:

1. Backward-to-forward voltage wave ratio at a particular port when all other ports have no input
2. Backward wave at a port due to forward wave at another port

Because we are dealing with sinusoidal voltages, the S parameters are usually complex quantities. A generalized S parameter concept drawing is shown in Fig. 20.13. For a two-port network, the S parameters have familiar common names:

- $S_{11}$  input reflection coefficient
- $S_{12}$  reverse insertion loss
- $S_{21}$  forward insertion loss
- $S_{22}$  output reflection coefficient

**example 20.14** For a two-port network, solve for the S parameters.



**solution** Our hypothesis is that the backward wave at any port is a linear combination of all incident waves, at least over some limited range of operation.

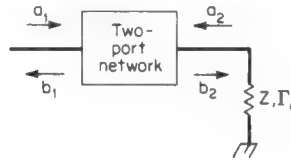
$$b_1 = S_{11}a_1 + S_{12}a_2 \quad b_2 = S_{21}a_1 + S_{22}a_2$$

If we terminate port 2 in a matched load, then  $a_2$  is 0. Thus

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}(0) & S_{11} &= \left. \frac{b_1}{a_1} \right|_{a_2=0} & S_{21} &= \left. \frac{b_2}{a_1} \right|_{a_2=0} \\ b_2 &= S_{21}a_1 + S_{22}(0) \end{aligned}$$

Similarly, 
$$S_{12} = \left. \frac{b_1}{a_2} \right|_{a_1=0} \quad S_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0}$$

**example 20.15** When  $a_2$  is not equal to 0, what is the input reflection coefficient  $b_1/a_1$ ? (This is the case if port 2 is terminated in a nonmatched load.)



**solution**

$$Z \neq Z_0 \quad \therefore \Gamma_v \neq 0$$

and

$$b_1 = S_{11}a_1 + S_{12}a_2$$

but

$$a_2 = \Gamma_v b_2 = \Gamma_v (S_{21}a_1 + S_{22}a_2)$$

Solving for  $a_2$  in terms of  $a_1$  yields

$$a_2 = \frac{\Gamma_v S_{21} a_1}{1 - \Gamma_v S_{22}} a_1$$

Substituting gives

$$b_1 = S_{11}a_1 + S_{12} \frac{\Gamma_v S_{21} a_1}{1 - \Gamma_v S_{22}} a_1$$

Finally,

$$\frac{b_1}{a_1} = S_{11} + \frac{S_{12} S_{21} \Gamma_v}{1 - \Gamma_v S_{22}}$$

**SKIN EFFECT** Loss of signal power in a microwave transmission circuit is a very important consideration. A coaxial line has its loss expressed in units of decibels per 100 ft. The amount of this loss is a function of the materials used in manufacturing the cable and is related also to the frequency of operation. Here we depart again from ordinary circuit theory in that resistance (loss) depends on frequency, which has no basis in conventional theory. The explanation of this phenomenon is called the *skin effect*. When signals of high frequency pass through transmission line, the currents do not penetrate the entire metallic wires but travel near the surface of the wires. This results in an apparent increase of resistance as the frequency is raised and the current-carrying area decreases. For example, coils in high-power transmitters sometimes are made of hollow tubing which is silver-coated. The hollow interior does not increase the resistance of the tubing for that frequency of operation. Often the interior of copper waveguide is flashed with silver since most of the energy barely penetrates the surface. A detailed explanation would require extensive use of electromagnetic theory.

Current decays within a conductor in an exponential manner. The skin depth is defined as that thickness in which the current is reduced by a factor  $1/e$  (note: this is analogous to the time constant). The skin depth  $\delta$  is related to frequency  $f$ , dc conductivity  $\sigma$ , and magnetic permeability  $\mu$ . For nonferrous metals, the value of  $\mu$  is the same as the free-space value  $\mu_0$ .

**example 20.16** Compare the skin depth at 60 Hz and 6 GHz in copper.



solution

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} = \frac{1}{\sqrt{\pi(60)(4\pi \times 10^{-7})(58 \times 10^6)}}$$

$$\delta \text{ (at 60 Hz)} = 8.53 \times 10^{-9} \text{ m} = 0.336 \text{ in}$$

$$\delta \text{ (at 6 GHz)} = 33.6 \text{ } \mu\text{in}$$

### 20.3 CIRCUIT COMPONENTS

In this section we describe many of the most frequently encountered microwave components. A few large categories are established first to arrange the components in a systematic manner. Active versus passive components are a first distinction. By *active* we mean any component that requires energy to function other than that supplied by the microwave signal itself. Otherwise, a component is passive. As an example, an amplifier is considered active because it requires dc input power to function, whereas an attenuator is a passive component, in general. We include in the active category components such as diode switches which do not produce more output power than is inputted but nonetheless require an external power source.

The next major subdivision is whether a component is linear or nonlinear. Linearity here implies the same meaning as in ordinary circuit theory: (1) the output signal is in direct proportion to the input signal, and (2) a combination of input signals will produce the same combination of responses at the output in the same proportions with no new signals created.

Further distinctions can be made based on the number of ports a component possesses. In the group of components having two or more ports, a further distinction is whether that component is reciprocal. A reciprocal two-port network would measure the same insertion loss and insertion phase regardless of the direction of the microwave energy. Clearly an amplifier is a nonreciprocal two-port device. During the era of microwave technology over the past 50 years, an enormous quantity of components have been developed. Many generic types of components exist simultaneously in different technologies. For example, a coupler may be made as a coaxial component or a waveguide component; similarly, an amplifier may be a traveling wave tube which is a thermionic device or a solid-state GaAs FET amplifier. In the family tree of Fig. 20.14, the components we will be discussing are arranged in their appropriate positions.

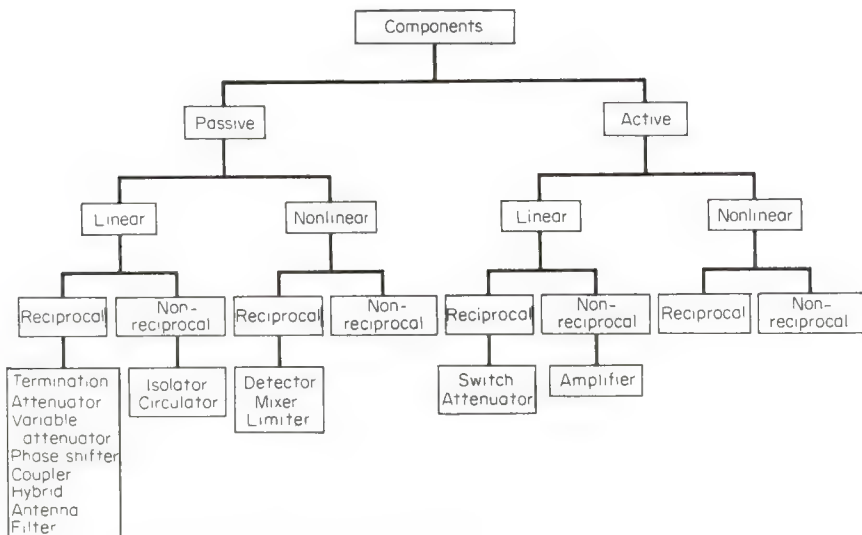


Fig. 20.14 Family tree of microwave circuit components.



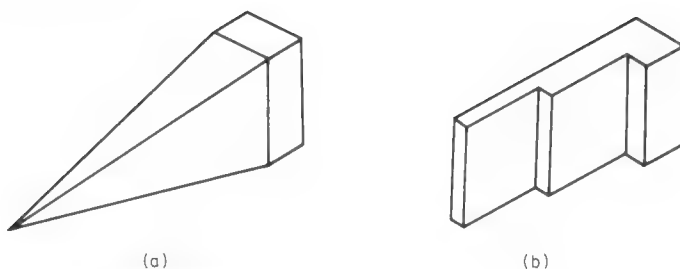


Fig. 20.15 Waveguide terminations.

## PASSIVE LINEAR RECIPROCAL COMPONENTS

**Terminations** Terminations are merely resistors designed to perform at microwave frequencies. They usually take a physical shape that is compatible to the transmission medium being used. Ordinary resistors with leads are not generally suitable for high-frequency application. The most frequently used terminations are  $50\ \Omega$  which correspond to the most widely used coaxial line impedance.

In waveguide components, the termination fits into the waveguide opening. It is made from a block of graphite and, as shown in Fig. 20.15, is shaped as a pyramid or a staircase. In all cases, the purpose of a "matched" termination is to provide a reflectionless ending for a transmission line. For certain laboratory calibration purposes, deliberately mismatched terminations are sold. The termination is a one-port network.

**Attenuator** The purpose of an attenuator is to reduce the magnitude of a signal. Unlike the potentiometer in ordinary circuit analysis, the signal reduction must occur while a matched impedance is maintained at input and output ports. Early attenuators were designed following the principles of the well-known T pad and Pi pad, as shown in Fig. 20.16. Because there are three resistor values to manipulate, one can see that it is possible to provide signal reduction and maintain a  $50\text{-}\Omega$  input and output impedance.

In waveguide systems, an attenuator takes the form of a dissipative (carbonized) card which is partially inserted in the waveguide opening. Many specialized attenuators have been developed for high-power applications where the dissipation of heat is the primary design difficulty.

**Variable attenuator** Various clever mechanizations for mechanically adjustable attenuators have been devised. More recently, voltage-controlled variable attenuators that have no moving parts have made large inroads, and we will discuss these devices later.

The stepped variable attenuator is simply a selector switch with a number of fixed atten-

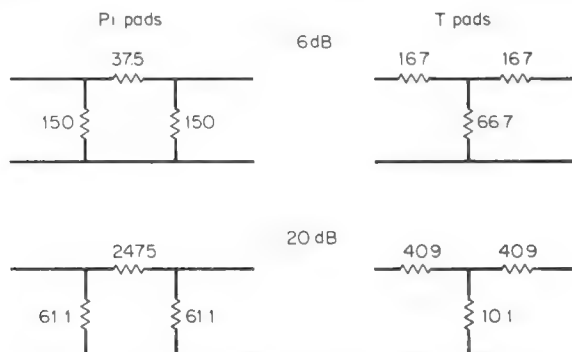


Fig. 20.16 Examples of Pi and T pads which attenuate the stated value and present a  $50\text{-}\Omega$  impedance to the source when the output is terminated in  $50\ \Omega$ .

uators chosen individually by rotating the switch. This device finds wide use at the lower microwave frequencies, i.e., below 2 GHz.

The continuously variable attenuator is made by ganging several potentiometers such that they produce a matched T or Pi pad at all positions of rotation. These devices also are best suited to the lower microwave frequencies.

Continuously variable attenuators of strip-line or microstrip construction are suited for the higher microwave frequencies. See Fig. 20.17. Typically a dissipative element is slid, by means of a lead screw, across the conductor of the transmission line. As the absorptive material approaches the line, more energy is removed from the transmission path.

Waveguide variable attenuators control insertion loss by adjusting the position of an absorbing card within the waveguide opening. See Fig. 20.18. Several standard methods are frequently used:

1. Through a slit in the broad wall of a waveguide a card is gradually inserted to vary the attenuation.
2. The card moves from the short wall of the waveguide gradually to the center of the waveguide for maximum attenuation.

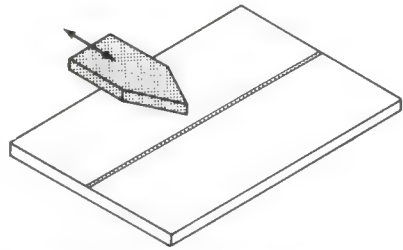


Fig. 20.17 Microstrip variable attenuator.

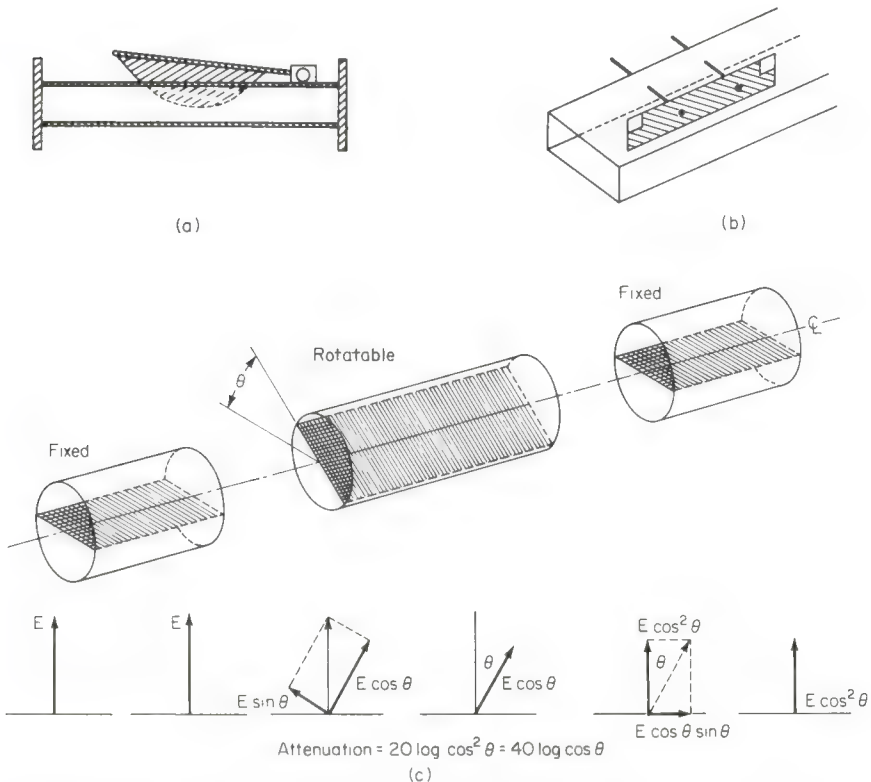


Fig. 20.18 Waveguide attenuators: (a) Flap attenuator. (b) Glass-vane attenuator. (c) Functional drawing including operating principle of the rotary-vane type of attenuator.

3. An absorbing card is rotated in a circular waveguide where its inclination relative to the electric field precisely determines the insertion loss. The rotary-vane attenuator is an extremely precise type of mechanical attenuator favored for use as a laboratory grade instrument.

**Phase shifter** The phase shifter is a device used to vary the phase of a signal passing through it in a nearly lossless process. Most simply, a phase shifter may be a coaxial line manufactured in such a manner that it can be lengthened physically. A stripline phase shifter is made by sliding a dielectric card over the conductor. The presence of the dielectric causes the velocity of propagation to decrease and thus increases the number of electric degrees of phase shift per linear dimension. More recently, semiconductor diodes are employed. Waveguide phase shifters are made very similarly to waveguide attenuators except that the absorbing material is replaced with dielectric in the case of side wall or top wall phase shifters. The rotary-vane phase shifter, like the rotary-vane attenuator, is a precise laboratory instrument.

**Coupler** The coupler is usually a three- or four-port device. Its function is to divide a signal entering a port into a fixed proportion among the other ports. This function can be accomplished with very small loss to the total power. All couplers of the low-loss type possess directional properties; for example, in Fig. 20.19a, the signal entering at port 1 produces 90 percent of its power at port 2 and 10 percent at port 3. However, a signal entering at port 2 produces 90 percent of its power at port 1 and ideally nothing at port 3. This type of coupler, as are all lossless couplers, is actually a four-port device with the fourth port terminated internally. Indeed, some couplers are sold as four-port couplers, as shown in Fig. 20.19b. A signal entering port 1 couples to port 3; a signal entering port 2 couples to port 4. None of a signal entering at port 2 should emerge at port 3; however, in real devices, some small amount does. That fraction is called *isolation*. The difference between isolation and coupling values is called *directivity*. Frequently, directional couplers are available in waveguide or coaxial versions with standard coupling values of 6, 10, 20, and 40 dB.

**Hybrids** Hybrids are, generally speaking, specialized 3-dB four-port couplers. Because they, like couplers, operate by proper phasing of signals, they are virtually lossless. The two major divisions of hybrids are based on the number of electric degrees which the coupler shifts the signals passing through it. The  $90^\circ$  hybrid, sometimes called the *short-slot hybrid* or *quadrature hybrid*, is shown in Fig. 20.20a with the phase relations. The  $180^\circ$  hybrid, often called the *magic T*, is shown in Fig. 20.20b. These devices are enormously useful for applications where signals need to be combined or canceled.

**Antennas** The antenna can be thought of as a device for matching the transmission line impedance to the impedance of free space. In that sense it is merely a special type of transmission line. In Sec. 20.4 the subject is discussed more fully.

**Filters** Filters are devices for selecting frequencies from all that are present; some are passed while others are attenuated. Most filters are constructed in a symmetric fashion and so are reciprocal. Because most filters contain only reactive and lossy elements, they are passive.

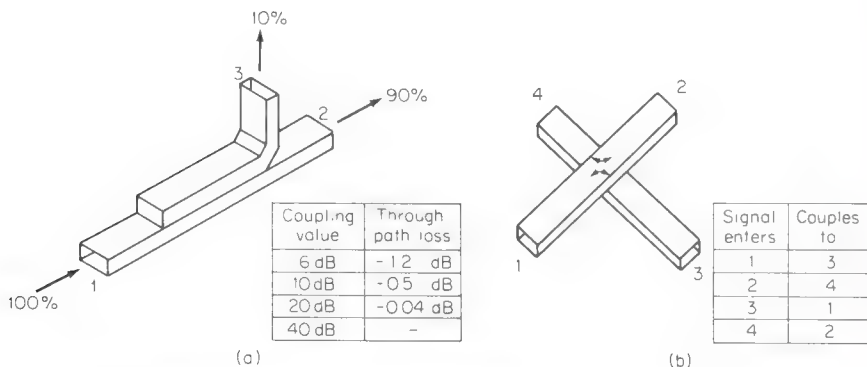
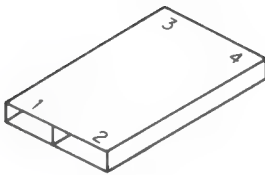


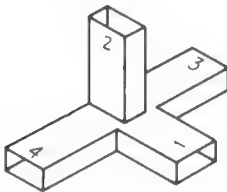
Fig. 20.19 (a) Broadwall coupler. (b) Crossguide coupler.



(a)

Response  
at port  
(mag/phase)

90° Hybrid Signal enters port				
	1	2	3	4
1	-	0	-3dB/0°	-3dB/-90°
2	0	-	-3dB/-90°	-3dB/0°
3	-3dB/0°	-3dB/-90°	-	0
4	-3dB/-90°	-3dB/0°	0	-



(b)

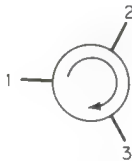
Response  
at port  
(mag/phase)

180° Hybrid Signal enters port				
	1	2	3	4
1	-	0	-3dB/0°	-3dB/0°
2	0	-	-3dB/0°	-3dB/180°
3	-3dB/0°	-3dB/0°	-	0
4	-3dB/0°	-3dB/180°	0	-

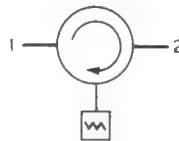
**Fig. 20.20** Hybrid 3-dB four-port couplers.

## PASSIVE LINEAR NONRECIPROCAL COMPONENTS

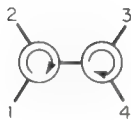
**Isolator** An isolator is a two-port device that allows a microwave signal to pass in one direction only. A small loss is incurred by the signal in the forward direction, and a substantial loss occurs in the reverse direction. This behavior is somewhat analogous to the action of a diode as it relates to the passage of current. The mechanism which produces this effect depends on the interaction of the microwave signal with a static magnetic field. Typical modern isolators at X band have insertion losses of 0.7 dB and reverse insertion loss of 20 dB or more. Both coaxial and waveguide versions are available over the entire microwave band.



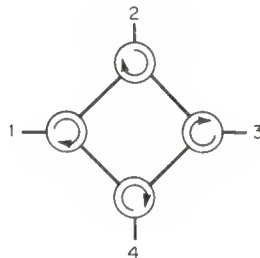
(a)



(b)



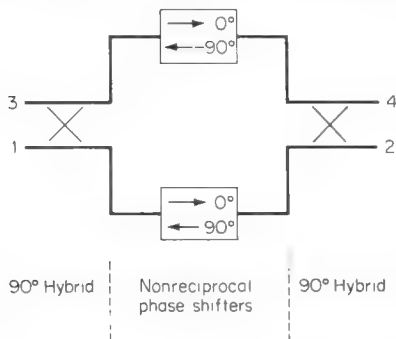
(c)



(d)

**Fig. 20.21** (a) Basic three-port circulator. (b) Circulator connected as an isolator. (c) Four-port circulator, two junctions. (d) Four-port circulator, four junctions.

**Circulator** The nonreciprocal interaction of a microwave signal with a static magnetic field results in passive devices which sort signals based on their direction of travel. The three-port circulator in Fig. 20.21a is probably the most fundamental of these devices since all others can be derived from it. A signal entering at port 1 emerges at port 2, a signal entering at port 2 emerges at port 3, and a signal entering port 3 emerges at port 1. The circulator with one port terminated becomes an isolator (Fig. 20.21b). When the magnetic field is supplied by a reversible electromagnet, a circulator becomes an electromagnetic switch.



**Fig. 20.22** One type of differential phase shift, four-port circulator.

When more than three ports are desired in the circulation loop, arrangements such as those shown in Fig. 20.21c and d are often used with the three-port circulator as the basic building block.

A widely used circulator mechanization is called the *Y junction*, which is available in coaxial and waveguide models. It is simple to build, may be made by using printed-circuit techniques, and is relatively inexpensive and physically compact. If exceptionally high power is to be handled, a device known as a differential phase shift circulator is often used. See Fig. 20.22. This device contains no circulators at all, but rather is made up of nonreciprocal phase shifters and hybrids. Another type of circulator worth mentioning is the waveguide Faraday rotational circulator. This device separates signals

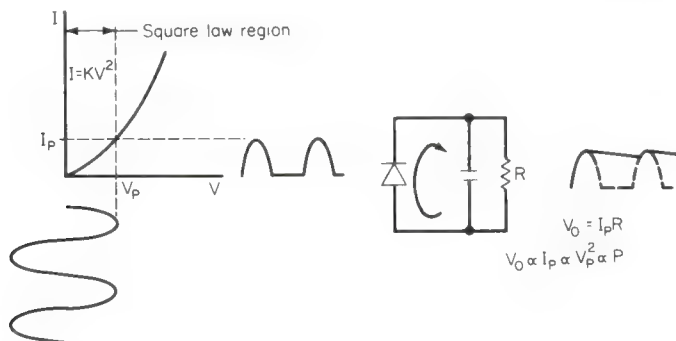
traveling in opposite directions by rotating them through  $45^\circ$  each as they interact with the magnetic field, but in the opposite sense of rotation.

## PASSIVE NONLINEAR RECIPROCAL COMPONENTS

**Detector** The detector is a junction diode. Its tiny size and low capacity enable it to probe a microwave signal and rectify it. Rectified voltage is then applied across a resistor, which completes the circuit. The small voltage that appears across the resistor usually must be amplified greatly in order to produce a useful signal. Today, in addition to p-n junction diodes, Schottky diodes and back diodes, which have a higher ratio of voltage produced to power input, are available. Within limited ranges of operation detectors are considered "square law." This simply means that the voltage produced is in direct proportion to the power incident on the diode, as shown in Fig. 20.23.

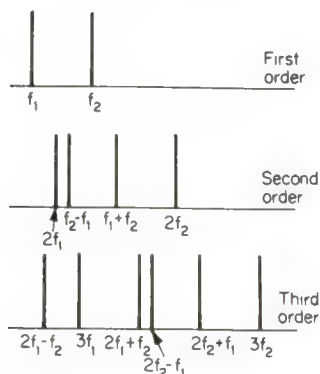
**Mixer** A mixer is one or more semiconductor junctions utilized to combine signals in a nonlinear fashion. Because the  $V$ - $I$  characteristic of a diode can be thought to contain high-order polynomial terms (Taylor series), two important results occur:

1. A single-frequency signal will produce its harmonics at the mixer output.



**Fig. 20.23** Square-law operation of a detector diode.

$$i(t) = K_0 + K_1 v + K_2 v^2 + K_3 v^3 + \dots$$



Single-tone intermodulation distortion

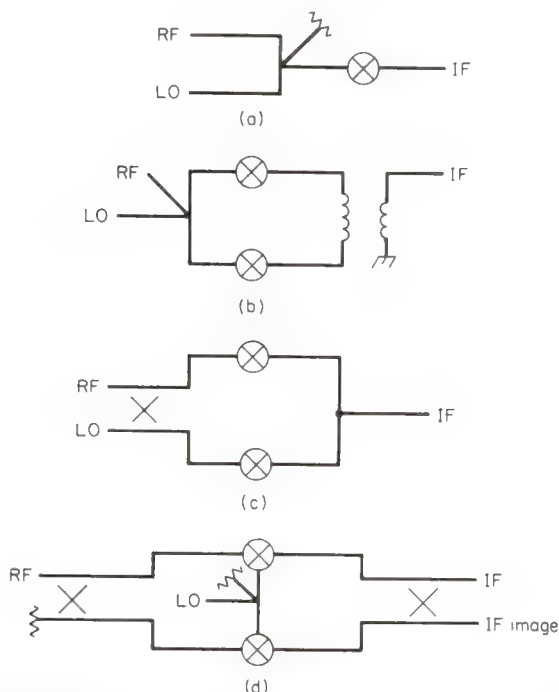
Harmonics of $f_R$	5	4	3	2	1	0	Harmonics of $f_L$
	82>90>90 82>90>90	80 83 87 80 82 90	78>90 90 80>90>90	72 78 74 75 78 82	80>90>90 83>90>90	67 74 71 73 75 78	5
	89>90>90 89>90>90	86>90>90 82>90>90	90>90>90 90>90>90	86>90>90 82>90>90	95>90>90 90>90>90	82>90>90 80>90>90	4
	58 65 69 60 68 72	56 62 74 58 62 76	63 68 76 65 71 78	54 54 57 56 58 61	68 72 76 68 72 78	50 53 53 54 59 59	3
	71 76 77 71 78 80	65 78 76 64 77 78	74 80 80 74 83 82	67 75 77 65 75 78	74 85 82 74 86 82	65 75 70 63 74 71	2
	16 24 30 16 24 30	0 0 0 0 0 0	28 40 40 26 38 40	11 13 13 10 11 11	34 43 44 32 39 44	22 28 25 17 20 20	1
	A B C 24 32 30	26 34 32 30 33 32	34 37 36 37 44 42	37 47 44 37 44 42	45 57 60 41 48 50	36 52 51 34 46 51	0

A: 0.05-200 MHz Class I mixer  
 B: 5-500 MHz Class I mixer  
 C: 2-500 MHz Class I mixer with baluns

$f_R = 49 \text{ MHz} @ -10 \text{ dBm}$   
 $f_L = 50 \text{ MHz}$

$f_L @ +10 \text{ dBm}$   
 $f_L @ +7 \text{ dBm}$

**Fig. 20.24** Mixer cross-product analysis. (Courtesy Watkins-Johnson, Palo Alto, Calif.)



**Fig. 20.25** (a) Single-ended mixer. (b) 180° balanced mixer. (c) 90° balanced mixer. (d) Image reject mixer.



**TABLE 20.4 Mixer Comparison Guide** (Courtesy **MACOM Microwave Associates, Burlington, Mass.**)

	Mixer comparison guide					
	Single-ended	Balanced (90°)	Balanced (180°)	Double-balanced	Image reject	Image recovery
Conversion loss	Good	Good	Good	Very good	Good	Excellent
VSWR LO, RF	Good, poor	Good, good	Fair, fair	Poor, poor	Good, good	Good, good
LO/RF isolation	Fair	Poor	Very good	Very good	Good	Very good
LO power required	+13	+5	+3	+10	+7	+7
Spurious rejection	Poor	Fair	Fair odd: fair	Good	Fair	Fair
Harmonic suppression	Poor	Fair	Even: good	Very good	Even: good Odd: fair	Even: good Odd: Fair
Third-order intercept	—	+13 dB <sub>m</sub>	+13	+18 dB <sub>m</sub>	+15 dB <sub>m</sub>	+15 dB <sub>m</sub>

2. Two different frequency signals will produce sum and difference frequencies at the mixer output and sums and differences of harmonics. See Fig. 20.24, which shows the amplitude of the cross products in decibels relative to the desired term,  $f_R \pm f_L$ , for three typical double balanced mixers and two bias levels.

The frequency conversion made possible by mixers enables the building of the superheterodyne receiver, which is the mainstay of radio, TV, and radar.

When individual mixer diodes are combined with other passive components, namely hybrids, they are given special names. These new configurations possess features such as suppression of the second harmonic of one or both inputs or rejection of an undesired image which are not present in the basic single-diode mixer. See Fig. 20.25 and Table 20.4. Although mixer diodes resemble detectors in appearance, it must be remembered that the mixer deals only with RF signals and no detection occurs. The mixer's three ports are often referred to as:

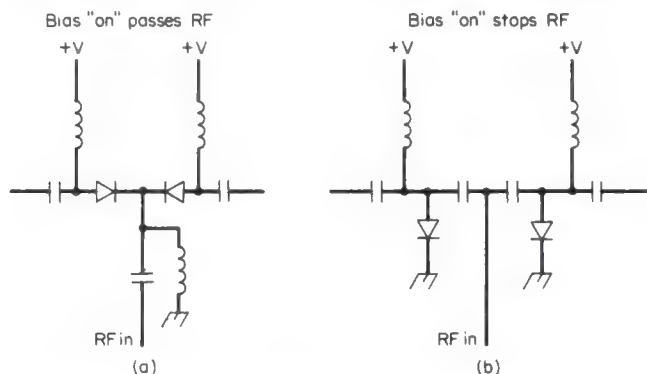
RF    radio frequency

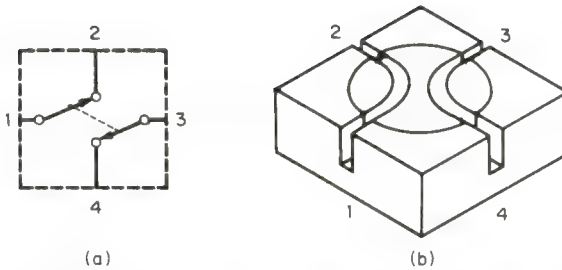
LO    local oscillator

IF    intermediate frequency

It is important that the semiconductor junction be operated in the proper nonlinear region of the  $V$ - $I$  curve; this is called *biasing* the mixer. Bias is supplied by local oscillator signal power and/or dc bias.

**Limiter** A limiter restricts the amount of power passing through it. Before the power reaches the limit value, the limiter has low loss. Two major types of limiters exist today: the diode limiter and the gas-tube limiter. A short-circuited diode placed in the transmission path will begin to conduct a current by detector action as the incident power is increased. The current that flows appears as a short circuit to the incident signal and reflects it. The gas tube is a short piece of waveguide filled with an inert gas. When the power level is reached that produces ionization of the enclosed gas, the plasma reflects the incident signal.

**Fig. 20.26** (a) Series diode SPDT switch. (b) Shunt diode SPDT switch.



**Fig. 20.27** E plane waveguide transfer switch. (a) Schematic diagram. (b) Cutaway view

### ACTIVE LINEAR RECIPROCAL COMPONENTS

**Switch** The application of miniature semiconductor diodes as microwave circuit elements has enabled a profusion of active control devices. A reverse-bias p-n junction between the inner and outer conductors of a coaxial cable would nearly appear to be an open circuit. When forward-bias current from an external source is passed through the diode, it appears as a short circuit to an incident signal. Switches fabricated as shown in Fig. 20.26 that use diodes would allow extremely rapid switching between ports. Both series diode and shunt diode switches are possible. In general, a switch containing multiple diodes produces an isolation (off state) of 20 dB per junction. Switches using diodes as the control element are made in all transmission media. Usually an electronic circuit accompanies the device and drives current through the diodes in response to external logic commands. The bias current is much greater than any detected RF current so the RF signal has minimal effect on diode bias.

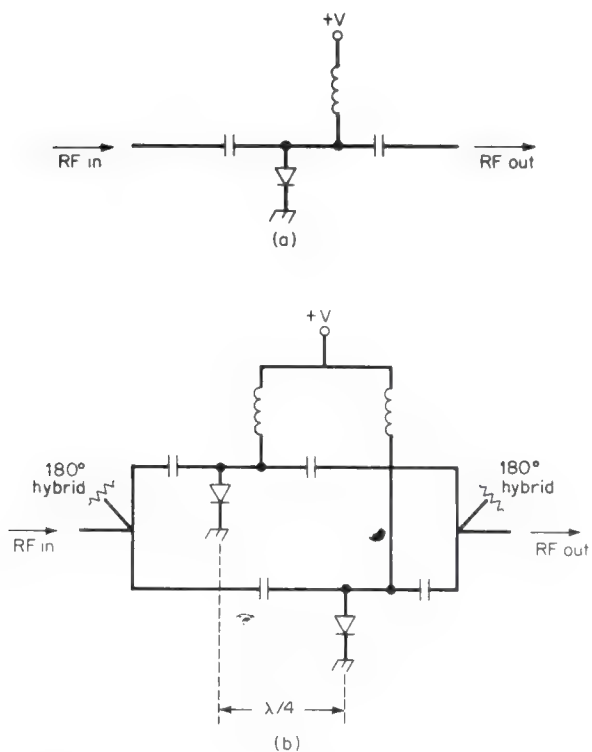
Mechanical coaxial and waveguide switches have long been used. Even those that have been motorized are extremely slow in comparison to the diode switch; however, insertion loss and isolation characteristics are superior. A type of mechanical switch frequently encountered in microwave work is called the *transfer switch*, as shown in Fig. 20.27. This type of switch is seldom encountered in electronic work.

**Attenuators** The diode voltage-controlled attenuator is similar to the switch in its operation. Rather than a bilevel control of diode bias, as in the diode switch, a continuous control of bias current is utilized. This mode of operation permits the insertion loss of the device to be varied over a wide range. Multiple-diode attenuators are frequently produced. Often an electronic circuit that modifies the control-voltage-to-diode-current transfer function precedes the diode so that the insertion loss in decibels is linearly proportional to control voltage. Two major types of diode attenuators are currently used: the reflective and absorptive types. They differ in how the energy not passed is dissipated. The reflective type has a higher input VSWR because the energy off the diode goes directly backward. In the absorptive type, phase relationships are such that the reflected energy is directed to a load. See Fig. 20.28.

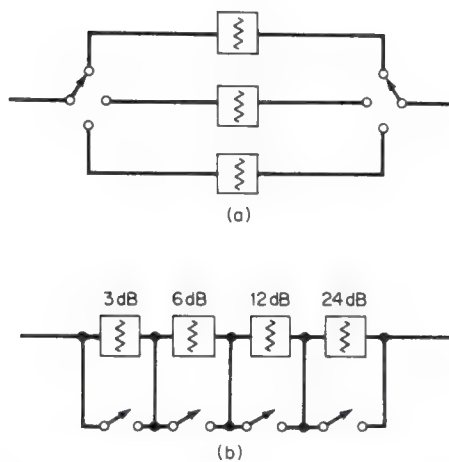
Another class of electronic attenuator consists of fixed pads and diode multithrow switches. This type is used when accuracy of attenuation and flatness of frequency response are desired. Fixed pads and diode switches also could be arranged in a binary circuit. See Fig. 20.29.

### ACTIVE LINEAR NONRECIPROCAL COMPONENTS

**Amplifiers** Today, nearly all low- to medium-power amplifiers use solid-state devices. For frequencies below 2 GHz, the bipolar transistor is the main device; at frequencies above 2 GHz, the gallium arsenide FET takes over. For amplifiers whose output power exceeds 1 W, thermionic devices, such as traveling wave tubes and crossed field amplifiers, dominate. The manufacture of solid-state microwave amplifiers is a highly competitive field and requires extremely fine photolithographic techniques at the higher frequencies. A "single-device" amplifier usually produces an 8-dB gain. Modern amplifiers do not require tuning and are inherently broadband. It is presently possible to purchase an octave band (8 to 16



**Fig. 20.28** (a) Reflective-type attenuator. (b) Absorptive-type attenuator.



**Fig. 20.29** (a) Multipad step attenuator. (b) 0- to 45-dB binary attenuator.

**TABLE 20.5 Low Noise Wideband Amplifiers**  
**Guaranteed Specifications at 25°C Case Temperature (Courtesy Avantek, Santa Clara, Calif.)**

Model	Frequency response (GHz)	Gain (dB)	Noise figure (dB)	Power output for 1-dB gain compression (dBm)		Gain flatness ( $\pm$ dB)	Typical intercept point for IM products (dBm)	VSWR (50 $\Omega$ ) maximum		Input power		Case drawing
	minimum	minimum	maximum	minimum	maximum	maximum	(dBm)	In	Out	Volts DC	Current (mA)	
<b>4 to 8 GHz</b>												
AMT-9032	4-8	17	4.5	+15	1.0		+25	2.0	2.0	+12	100	IC2
AMT-9033	4-8	27	4.5	+15	1.0		+25	2.0	2.0	+12	150	IC4
AMT-9034	4-8	36	4.5	+15	1.5		+25	2.0	2.0	+12	200	IC4
AMT-9035	4-8	45	4.5	+15	2.0		+25	2.0	2.0	+12	250	IC6
<b>5 to 10 GHz</b>												
AMT-9052	4-8	10	8.5	+20	0.75		+30	2.0	2.0	+12	220	IC2
AMT-9053	4-8	19	6.0	+20	1.0		+30	2.0	2.0	+12	270	IC4
AMT-9054	4-8	28	5.0	+20	1.0		+30	2.0	2.0	+12	320	IC4
AMT-9055	4-8	37	4.5	+20	1.5		+30	2.0	2.0	+12	340	IC6
<b>5 to 10 GHz</b>												
AMT-10033	5-10	27	5.5	+13	1.0		+23	2.0	2.0	+12	160	IC4
AMT-10034	5-10	37	5.5	+13	1.5		+23	2.0	2.0	+12	210	IC4
AMT-10035	5-10	46	5.5	+13	2.0		+23	2.0	2.0	+12	270	IC6
<b>7 to 12 GHz</b>												
AMT-12033	7-12	22	4.5	+15	1.5		+25	2.0	2.0	+12	150	IX4
AMT-12034	7-12	30	4.0	+15	1.5		+25	2.0	2.0	+12	200	IX4
AMT-12035	7-12	37	4.0	+15	2.0		+25	2.0	2.0	+12	250	IX6
AMT-12036	7-12	44	4.0	+15	2.0		+25	2.0	2.0	+12	300	IX6
<b>8 to 12.4 GHz</b>												
AMT-12433	8-12.4	22	4.5	+15	1.5		+25	2.0	2.0	+12	150	IX4
AMT-12434	8-12.4	30	4.0	+15	1.5		+25	2.0	2.0	+12	200	IX4
AMT-12435	8-12.4	37	4.0	+15	2.0		+25	2.0	2.0	+12	250	IX6
AMT-12436	8-12.4	44	4.0	+15	2.0		+25	2.0	2.0	+12	300	IX6

TABLE 20.5 Low Noise Wideband Amplifiers (continued)

Model	Frequency response (GHz)	Gain (dB)		Noise figure (dB)	Power output for 1-dB gain compression (dBm)		Gain flatness (±dB) maximum	Typical intercept point for IM products (dBm)	VSWR (50 Ω)		Input power		Case drawing
		minimum	maximum		minimum	maximum			In	Out	Volts DC	Current (mA)	
6 to 18 GHz													
AWT-18633	6-18	16	7.5		+12		1.0	+22	2.0	2.0	+12	160	IX4
AWT-18634	6-18	22	7.5		+12		1.0	+22	2.0	2.0	+12	200	IX4
AWT-18635	6-18	28	7.5		+12		1.5	+22	2.0	2.0	+12	250	IX6
AWT-18636	6-18	34	7.5		+12		1.5	+22	2.0	2.0	+12	300	IX6
AWT-18637	6-18	40	7.5		+12		2.0	+22	2.0	2.0	+12	350	IX8
AWT-18638	6-18	46	7.5		+12		2.0	+22	2.0	2.0	+12	400	IX8
AWT-18656	6-18	34	7.5		+20		1.5	+28	2.0	2.0	+12	350	IX6
8 to 18 GHz													
AWT-18033	8-18	15	7.5		+13		2.0	+22	2.0	2.0	+12	150	IX4
AWT-18034	8-18	20	7.5		+13		2.0	+22	2.0	2.0	+12	200	IX4
AWT-18035	8-18	25	7.5		+13		2.0	+22	2.0	2.0	+12	250	IX6
AWT-18036	8-18	30	7.5		+13		2.0	+22	2.0	2.0	+12	300	IX6
AWT-18037	8-18	35	7.5		+13		2.5	+22	2.0	2.0	+12	350	IX8
AWT-18038	8-18	40	7.5		+13		2.5	+22	2.0	2.0	+12	400	IX8
AWT-18039	8-18	45	7.5		+13		2.5	+22	2.0	2.0	+12	450	IX10
AWT-18057	8-18	35	7.5		+20		2.0	+28	2.0	2.0	+12	350	IX8
12 to 18 GHz													
AMT-18033	12-18	16	7.0		+13		2.0	+20	2.0	2.0	+12	150	IX4
AMT-18034	12-18	22	7.0		+13		2.0	+20	2.0	2.0	+12	200	IX4
AMT-18035	12-18	27	7.0		+13		2.0	+20	2.0	2.0	+12	250	IX6
AMT-18036	12-18	33	7.0		+13		2.0	+20	2.0	2.0	+12	300	IX6
AMT-18037	12-18	38	7.0		+13		2.0	+20	2.0	2.0	+12	350	IX8
AMT-18038	12-18	44	7.0		+13		2.0	+20	2.0	2.0	+12	400	IX8

GHz) GaAs FET amplifier with 60-dB gain and a saturated output power of +20 dBm which is unconditionally stable. An example of the diversity of available amplifiers is shown in Table 20.5.

## 20.4 ANTENNAS

The subject of antennas is in no way restricted to the microwave region, but is frequently introduced during the study of microwaves because of the profusion of antenna types in that frequency region. Some antennas associated with microwaves have counterparts at the lower frequencies while others have commonality with optical devices. Because the microwave wavelengths are of a convenient size, being neither too large nor too small, a great variety of antenna types have been developed.

In this discussion we present an overview of antenna principles and types, but avoid detailed discussion of electromagnetic field theory. We accept without resorting to Maxwell's equations that radio energy can be made to propagate through space. The means by which energy transitions into space or is collected for reception can be a general definition of an antenna.

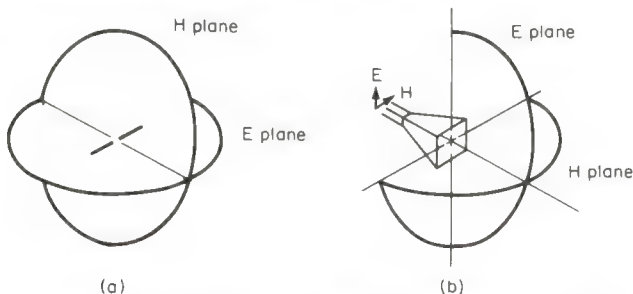
**RECIPROCITY** All antennas are reciprocal. We extend the definition of reciprocity at this time to include the directional properties of any antenna. For example, consider a dipole antenna connected to a radio transmitter. Current produced by the transmitter flows up and down the dipole antenna; energy radiates into space, and the signal strength is greatest broadside to the dipole. The principle of reciprocity now implies that when this antenna is used for receiving signals, those which emanate from a broadside direction will be received to a greater degree than those coming from other directions. A more succinct way of stating the same thing is that the radiation pattern of an antenna is identical on transmission or reception.

**ISOTROPIC RADIATOR** Before we define directivity and gain, the concept of isotropic radiation must be understood. The sun is an isotropic radiator of light, which is to say that the light intensity or the watts per square meter of light energy are directed equally in all directions. An isotropic antenna would produce concentric spheres of constant power intensity centered on the isotropic antenna. If one integrated the power intensity over a surface enclosing the antenna, one could calculate the total power radiated. This last fact is a consequence of the inverse-square law nature of propagating electromagnetic energy.

**ANTENNA PATTERNS** An antenna pattern is most often a plot of the relative power intensity at a fixed distance from the antenna under test as a function of angular direction. Antenna patterns are measured at distances sufficiently removed that one can ignore the effect of radial distance on the relative pattern shape. This is called the *far field pattern*. The line of demarcation is generally taken as

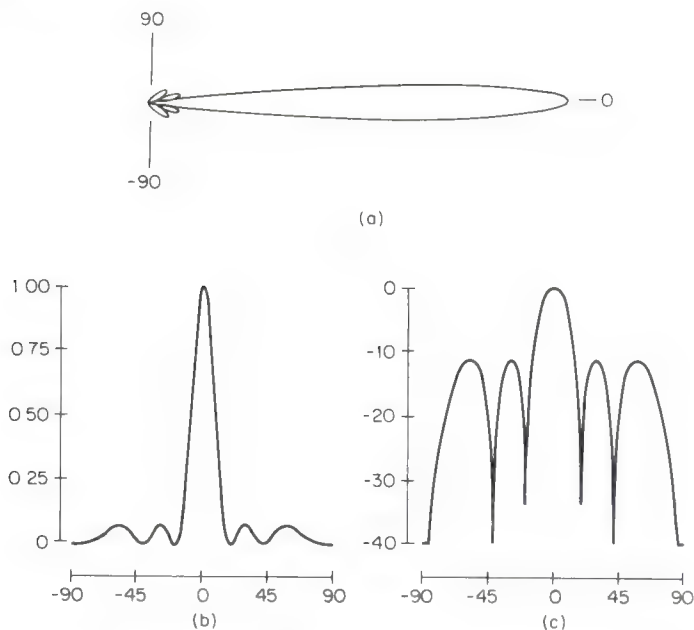
$$R = \frac{2L^2}{\lambda}$$

where  $R$  is the minimum radial distance to the point of observation,  $L$  is the maximum linear dimension of the antenna under test, and  $\lambda$  is the operating wavelength.



**Fig. 20.30** Principal measurement planes for (a) dipole antenna, (b) horn antenna.





**Fig. 20.31** Data formats for antenna patterns: (a) Direct power, polar plot. (b) Direct power, rectangular plot. (c) Logarithmic power (dB), rectangular plot.

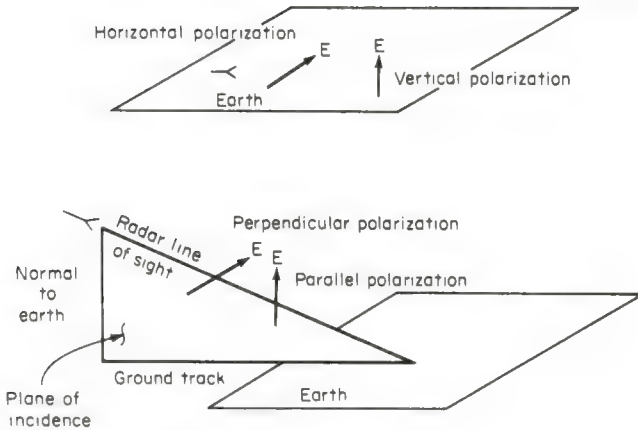
Because we are measuring a three-dimensional characteristic on two-dimensional paper, many patterns are required to determine the relative power intensity at all points in space (see Fig. 20.30). The plane parallel to the electric ( $E$ ) field and the plane parallel to the magnetic ( $H$ ) field are usually orthogonal and can be the principal planes of the antenna's radiation pattern. Sometimes the principal planes are determined relative to mechanical references as, for example, with circularly polarized antennas. The principal planes are equivalently the equatorial plane and the plane of the prime meridian. When antennas are measured relative to terrestrial references, azimuth and elevation are often the principal planes.

The data may be plotted in polar or rectangular form, and the amplitude may be in direct power or decibels. See Fig. 20.31.

**POLARIZATION** In describing linearly polarized antennas (the electric field is oriented in a constant direction), the plane parallel to the electric field is named for its orientation relative to a fixed reference. The antenna is then said to be polarized in that direction. Vertical and horizontal polarization refers to the direction of electric field relative to the earth. Parallel or perpendicular polarization refers to orientation relative to the plane of incidence, as in airborne radar sets. See Fig. 20.32.

**DIRECTIVITY** Antennas are useful because we can increase the power intensity in one direction and reduce it in another. The total power radiated, of course, remains unchanged. The ratio of the maximum power intensity for a given antenna to that radiation intensity of an isotropic radiator of the same total power is called *directivity*.

**GAIN** The gain of an antenna, while similar to directivity and often confused with it, has a different definition. *Gain* is the ratio of maximum power intensity to the maximum power intensity of a reference antenna having the same power input. Thus the gain characteristic includes losses between the antenna input terminals and free space. Recall that directivity makes no statement about the power input from a generator; thus the gain, relative to an isotropic radiator, of an antenna is less than its directivity.



**Fig. 20.32** Definition of polarization orientation.

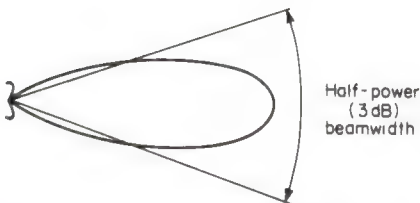
**EFFECTIVE AREA** Effective area is most easily visualized by thinking of antennas in the receiving mode. The antenna sits in a flux of energy, collects it, and delivers it to a load. If we assume the process is lossless, then the power in the load is the intensity of the incident energy multiplied by the effective area of the antenna. All antennas possess an effective area even though their physical areas are small (e.g., a wire dipole antenna). The effective area does not always have a physical significance except in cases of horns or reflector-type antennas where the effective area is 0.5 to 0.7 times the physical area. Antenna effective area is simply another formulation for directivity, and a known fixed relation applies:

$$D = \frac{4\pi A}{\lambda^2}$$

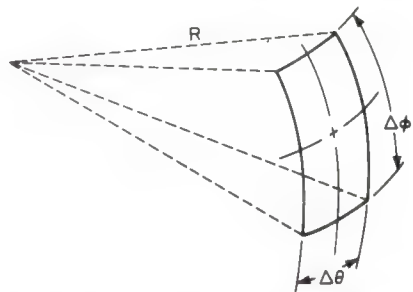
where  $D$  is directivity,  $A$  is effective area, and  $\lambda$  is operating wavelength.

**BEAMWIDTH** Many antennas are designed to concentrate the energy in one or two directions. A measure of that concentration is called *beamwidth*. The radiation intensity decreases continuously from its maximum value as one moves away from the peak of the pattern and does not abruptly stop at the beamwidth angles, as might be construed. Normally the angular difference between the points where the radiation intensity is one-half of or 3 dB below its maximum value is called the beamwidth. See Fig. 20.33. As the beam becomes narrower, the directivity increases since the same energy is squeezed into a narrower solid angle.

**ESTIMATING BEAMWIDTH AND DIRECTIVITY** The relation between beamwidth and directivity is now approximated. See Fig. 20.34. If one assumes all power is contained within



**Fig. 20.33** Beamwidth of an antenna pattern.



**Fig. 20.34** The principal plane beamwidths and solid-angle approximation.

the principal-plane beamwidths and is constant in that region and there is only one main lobe, then the power intensity over that area is the total power radiated divided by the area. Thus using radian angles

$$D = \frac{\text{maximum power intensity}}{\text{power intensity from isotropic source of equal total power}} \\ = \frac{W/A}{W/(4\pi R^2)} = \frac{W/(R \Delta\theta R \Delta\phi)}{W/(4\pi R^2)} = \frac{4\pi}{\Delta\theta \Delta\phi}$$

Converting to degrees yields

$$D = \frac{4\pi}{(\pi/180) \Delta\theta_1 (\pi/180) \Delta\phi_1}$$

or

$$D = \frac{41\,252}{\Delta\theta_1 \Delta\phi_1}$$

To make this formulation more correctly reflect real-world patterns, i.e., to account for the approximations made, the number in the numerator is reduced to about 30 000.

Another useful rule of thumb, though not derived from the above, is mentioned at this point. When the size and frequency of an antenna are known, the beamwidth (and thus directivity) can be calculated:

$$\theta = \frac{60}{L_\lambda}$$

where  $\theta$  is the half-power beamwidth in degrees and  $L_\lambda$  is the physical dimension, expressed as a number of wavelengths.

**example 20.17** It is desired to take antenna pattern measurements at 20 GHz on a 3-in by 5-in aperture horn. Determine the pertinent parameters.

**solution** The minimum distance for accurate measurements is

$$R = \frac{2L^2}{\lambda} = \frac{2(5)^2}{11.8/20} = \frac{50}{0.59} = 84.7 \text{ in}$$

The beamwidths are (interfering objects are kept clear of this region)

$$\theta = \frac{60}{5/0.59} = 7.08^\circ \quad \phi = \frac{60}{3/0.59} = 11.8^\circ$$

The estimated directivity is

$$D = \frac{30\,000}{7.08 \times 11.8} = 359, \text{ or } 25.5 \text{ dB}$$

At 84.7 in the size of the beam is

$$84.7 \times 7.08 \times \frac{\pi}{180} = 10.5 \text{ in} \\ 84.7 \times 11.8 \times \frac{\pi}{180} = 17.4 \text{ in}$$

**ANTENNA ARRAYS** The earliest technique used to shape the radiation pattern was to arrange antennas in geometric figures and control the phase angle and amplitude at each antenna so that the signals would add in phase in one direction and cancel at other directions. This technique is the basis of all antenna pattern analysis. The individual antennas making up the array are called *elements*. They may be vertical towers standing in a field for a broadcast station, slots cut periodically into a waveguide, or coplanar parallel metal rods, as in a TV antenna. These arrays are called *driven* arrays because each element is connected to the others and to the electronic device by a transmission line. In a *parasitic* array only one element is connected to the transmission line and the other elements are

excited by mutual coupling to the driven element. We discuss the principles of driven arrays.

**example 20.18** For two isotropic in-phase equal-amplitude radiators spaced 1 wavelength apart, find an expression for the far-field pattern.

**solution** The path-length difference to the point of observation depends on the separation and angle of observation

$$l = d \sin \theta$$

The electric phase difference at that observation point is

$$360 \frac{d}{\lambda} \sin \theta = \psi$$

In the figure the voltage from radiator 2 lags that from radiator 1 by  $\psi$  degrees. Summing the voltages at  $P$  yields

$$\begin{aligned} \tilde{V} &= \underset{(1)}{1/0} + \underset{(2)}{1/-\psi} \\ &= 1 + e^{-j\psi} \\ &= (e^{j\psi/2} + e^{-j\psi/2})e^{-j\psi/2} \\ &= \left(2 \cos \frac{\psi}{2}\right) (e^{-j\psi/2}) \end{aligned}$$

The magnitude is

$$|\tilde{V}| = 2 \cos \frac{\psi}{2}$$

The power is proportional to  $|\tilde{V}|^2$ , and normalizing for the number of elements yields

$$P = \cos^2 \frac{\psi}{2}$$

$$\begin{aligned} \text{Now } \psi &= 360 \frac{d}{\lambda} \sin \theta \quad \text{and} \quad d = \lambda \\ &= 360 \sin \theta \end{aligned}$$

$$\begin{aligned} \text{Finally,} \\ P &= \cos^2 (180 \sin \theta) \end{aligned}$$

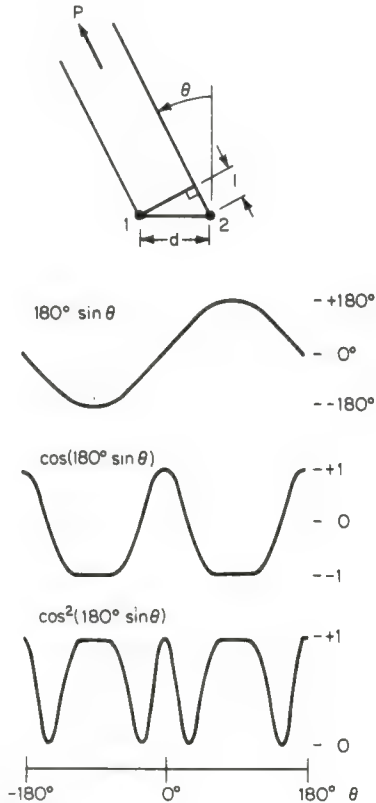
Notice that in Example 20.18 there are four lobes. The two at  $\theta = 0^\circ$  and  $\theta = 180^\circ$  are primary lobes, and the two at  $\theta = +90^\circ$  and  $\theta = -90^\circ$  are called *grating lobes*. Grating lobes cannot arise if separation between elements is less than approximately  $0.6\lambda$ .

A similar analysis on  $N$  equal-amplitude uniform phase progression isotropic radiators yields the useful expression

$$P = \left[ \frac{\sin (N\psi/2)}{N \sin (\psi/2)} \right]^2$$

where  $N$  is the number of elements,  $\psi = \phi + 360(d/\lambda) \sin \theta$ , and  $\phi$  is the feed-driven electrical phase-shift between adjacent elements. The case of two elements is explored in Fig. 20.35.

**PATTERN MULTIPLICATION** Sometimes the radiation patterns of complicated arrays can be generated by multiplying the array patterns of subdivisions of the original array by the pattern of isotropic radiators representing the overall array. As in Fig. 20.36, the subdivisions are called *element patterns*, and the overall representation is called an *array factor*. In Fig. 20.37, we demonstrate the binomial array by pattern multiplication. This array is said to have a taper; i.e., the central elements have greater amplitude than the edge elements.



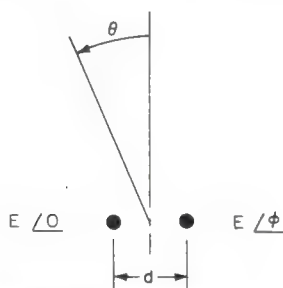
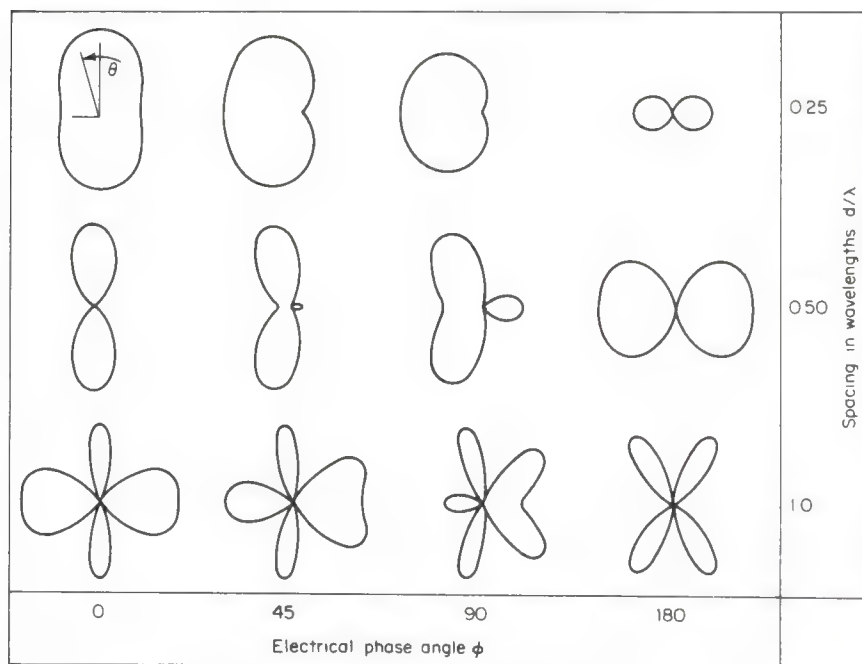
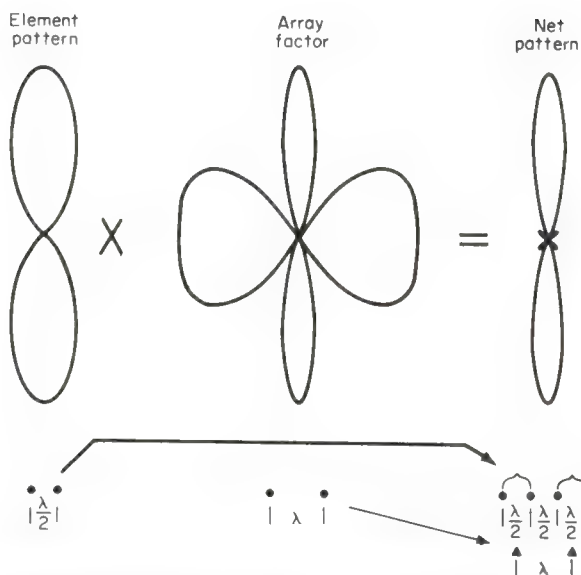


Fig. 20.35 Two equal-amplitude isotropic radiators.

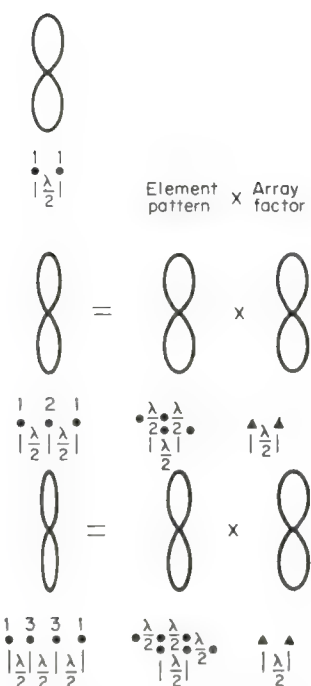
### MICROWAVE ANTENNAS

**Dipoles** Although dipoles are used throughout the frequency spectrum, they are important to microwave antennas despite their small physical size. Dipoles can be printed on circuit boards along with their feed networks to form planar arrays. They also could be printed in diminishing lengths to make a log-periodic antenna (see Fig. 20.38a). In the log-periodic antenna, the spacing between the  $N$ th and  $(N + 1)$ st element is a fixed fraction of a wavelength of the  $N$ th element. This is done so that each element when excited will "see" identical environments.

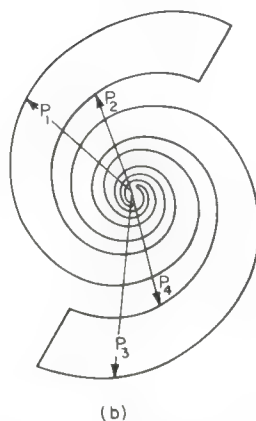
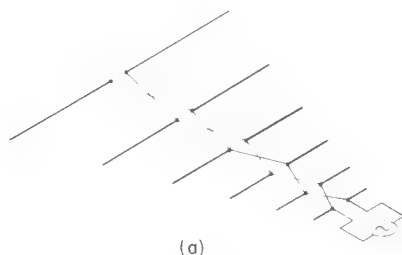
**Spirals** The equiangular spiral shown in Fig. 20.38b is a logarithmic structure. The radius of the spiral and the width of the element increase in proportion to the angle subtended. Low frequencies excite the large-radius elements, and high frequencies excite the smaller-radius portions. The spiral antenna is a convenient means of producing circular polarization. Spirals are, in general, broad-beamed and low in gain. Another spiral antenna frequently encountered is called the *Archimedes spiral antenna*. In it the element width and pitch are constant.



**Fig. 20.36** Pattern generation for a four-element in-phase, equal-amplitude array by the method of pattern multiplication.



**Fig. 20.37** Binomial array. Elements are spaced one-half wavelength apart, in phase, having amplitudes proportional to the binomial expansion coefficients. The overall pattern is derived by pattern multiplication in this case. The binomial array has no side lobes.



**Fig. 20.38** Types of broadband antennas: (a) Log-periodic dipole array. (b) Equiangular spiral.

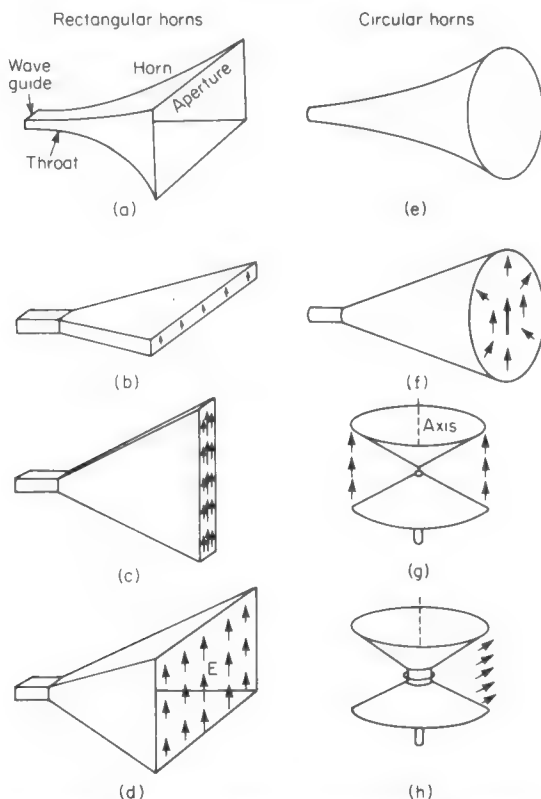


**Horns** A horn antenna is a waveguide that is flared in one or two dimensions. Figure 20.39 illustrates the common rectangular and circular types. The gradual flaring creates a plane wave at the aperture of the horn. The larger the aperture, the narrower the beam-width, and vice versa. The theory of horn antennas is so well understood that horns are often used as gain calibration standards. One can regard the emerging wavefront as a linear array of infinitely small point sources. See Fig. 20.40.

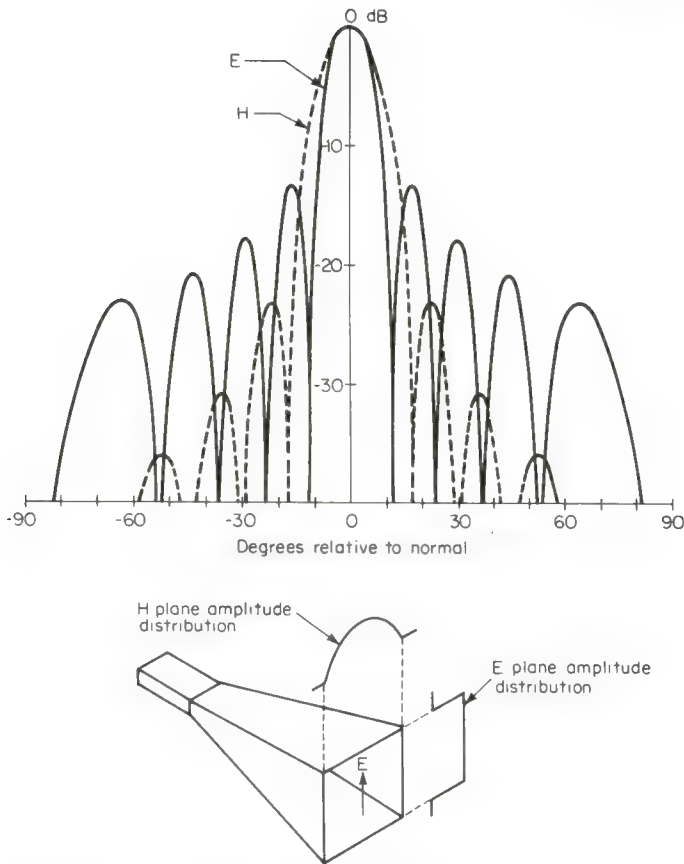
**Focusing antennas** The antennas discussed next operate on quasi-optical principles. The analogy between reflective telescopes and reflector antennas is fairly clear. A metal surface takes the place of a polished mirror. The dielectric lens has the common glass lens as its analogy. The idea of collimating or focusing the sun's rays into a point by using shaving mirrors or a magnifying glass is a common experience shared by many experimenters.

**Reflector Type.** It is possible to generate an aperture with an emerging plane wave by using the geometric properties of a parabola. A parabola is the locus of points equidistant from a point and a line. It follows that waves emitted from the focal point of a parabola of revolution or a cylindrical parabola will add in phase in a plane perpendicular to the parabolic axis, the aperture. See Fig. 20.41*a* and *b*.

**Lens Type.** The planar wavefront also may be created by using a lens. Spherical wavefronts emitted from a source located at the lens' focus are straightened out as shown in Fig. 20.41*c*. The center of the lens, being thicker, will delay the wavefronts more than at the edges of the lens. In Fig. 20.41*d*, metal slats parallel to the electric field accelerate the wavefronts at the edges of the lens so that they catch up to the central portion. This is possible because the phase velocity is greater in a waveguide than in free space.



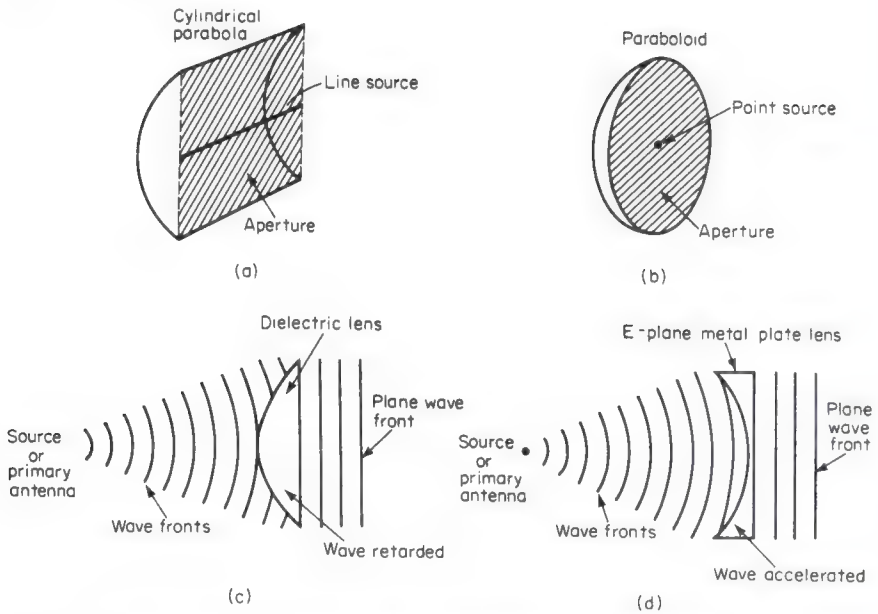
**Fig. 20.39** Types of rectangular- and circular-horn antennas.



**Fig. 20.40** Ideal *E* and *H* plane directivity patterns of a 5-wavelength square-horn antenna fed by a waveguide in the  $TE_{10}$  mode.

## 20.5 BIBLIOGRAPHY

- Brown, R. G., et al.: *Lines, Waves and Antennas*, Ronald, New York, 1961.  
 Jasik, Henry: *Antenna Engineering Handbook*, McGraw-Hill, New York, 1961.  
 Jordan, Edward C.: *Electromagnetic Waves and Radiating Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1950.  
 Kaufman, M., and A. Seidman: *Handbook of Electronics Calculations*, McGraw-Hill, New York, 1979.  
 Kraus, John D.: *Antennas*, McGraw-Hill, New York, 1950.  
 Lance, Algje L.: *Introduction to Microwave Theory and Measurements*, McGraw-Hill, New York, 1964.  
 Schelkunoff, S. A., and H. T. Friis: *Antennas—Theory and Practice*, Wiley, New York, 1952.



**Fig. 20.41** Types of focusing antennas: (a) and (b) Reflector types. (c) and (d) Lens types.

# Chapter 21

## Fiber-Optic Systems

**Joseph R. Cleveland**

*Senior Manager*

*Transmission Technology Facility, MCI*

*Richardson, Texas*

### 21.1 INTRODUCTION

Through the centuries people have been intrigued by the use of light to transmit information. Many quests have been initiated to realize optical communication systems which are practical and economical. One such example is the rapid progress made in fiber-optic technology. The large number of present and planned fiber-optic communication installations attests to how readily this technology has been adopted.

As with the established wire or radio communication technology, a basic fiber-optic system (Fig. 21.1) consists of a transmitter (light source with modulator or driver), propagation medium (optical fiber with connectors or splices), and a receiver (light detector and preamplifier). The light source may be either a light-emitting diode (LED), similar to that used in pocket calculators, or the more intense injection laser diode (ILD). Information is sent as intensity modulation (IM) of the light produced by the transmitter. In keeping with radio frequency (RF) technology, the optical signal is referred to as the *carrier*. Although frequency or phase modulation of the carrier can be accomplished, the technology is not yet available for practical implementation.

The position of the optical carrier in the RF spectrum is identified in Fig. 21.2. Note that optical communication frequencies range from 300 to 600 terahertz (THz). In the optical range it is common to specify the wavelength  $\lambda$ , given by

$$\lambda = \frac{c}{f} \quad (21.1)$$

where  $f$  is the frequency and  $c$  is the speed of light, which is  $3 \times 10^8$  m/s in vacuum. Three units of wavelength commonly used are

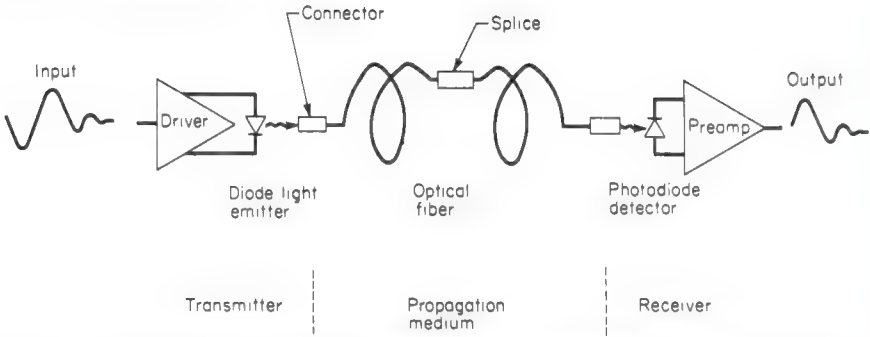
1. Angstrom ( $\text{\AA}$ ) =  $10^{-10}$  m
2. Nanometer (nm) =  $10^{-9}$  m
3. Micrometer ( $\mu\text{m}$ ) =  $10^{-6}$  m

**example 21.1** An optical carrier has a frequency of 400 THz. What is its wavelength?

**solution**

$$\begin{aligned} \lambda &= \frac{3 \times 10^8 \text{ m/s}}{400 \times 10^{12} \text{ Hz}} \\ &= 7.5 \times 10^{-7} \text{ m} \\ &= 0.75 \mu\text{m} = 750 \text{ nm} = 7500 \text{\AA} \end{aligned}$$

## 21-2 Fiber-Optic Systems

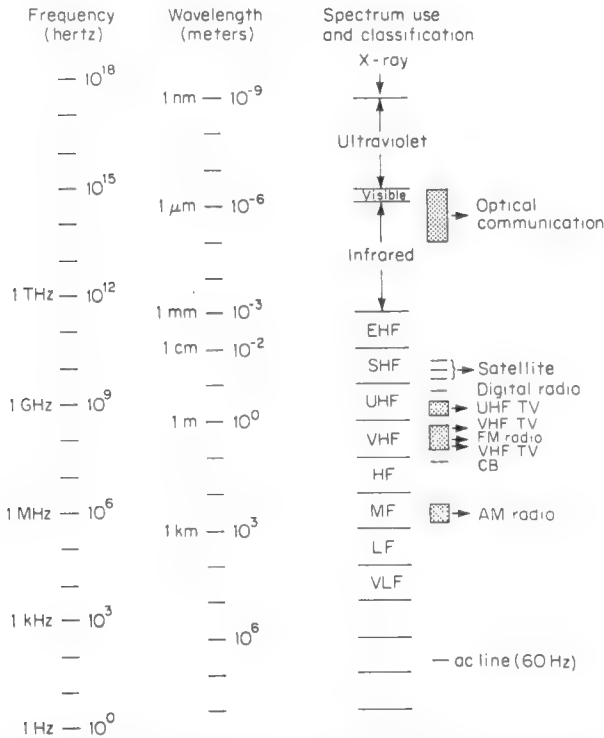


**Fig. 21.1** Block diagram of a basic fiber-optic link. A duplex link requires an identical system but in reverse order.

Presently, fiber-optic systems operate in the range of 0.6 to 1.6  $\mu\text{m}$ .

Fiber-optic systems offer a number of distinct advantages over wire or radio systems. Chief among these are

1. Low transmission losses
2. Greater bandwidths
3. Lower weight, less bulky cables



**Fig. 21.2** Electromagnetic wave spectrum. The RF spectrum lies below 300 GHz, and the optical spectrum lies between 300 GHz and 30 000 THz. Optical transmitters used in fiber-optic systems emit light in the visible and near-infrared portions of the spectrum between 200 and 600 THz.

4. Immunity to inductive interference
5. Excellent signal confinement

These advantages relate to communications technology as follows:

1. Low transmission loss means that the light signal in the optical guide suffers little loss between the transmitter and receiver. Fibers are available now with an attenuation rating as low as 0.3 decibel per kilometer (dB/km) (a power reduction of 6.7 percent for each 1-km length). In contrast to coaxial cable characteristics, fiber attenuation is independent of modulation frequency, as shown in Fig. 21.3.

2. A greater bandwidth means that more information can be conveyed than with wire or radio communication systems. This property coupled with low fiber attenuation provides long-distance, high-data-rate transmission links. For example, transmission of  $2 \times 10^9$  bits/s over a 44-km fiber link has been demonstrated.

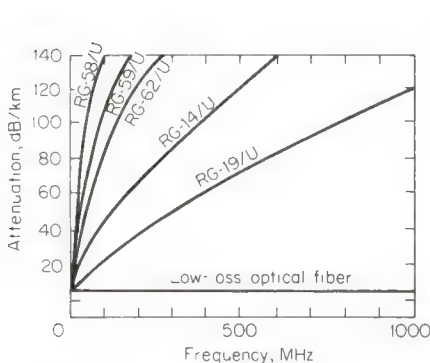
3. Lower weight and less bulky cables permit increased transmission capacity in underground conduits. A 7-mm-diameter optical cable can replace a 7.6-cm (3-in) diameter, 900-pair copper cable (see Fig. 21.4) and provide the same information bandwidth. This size reduction permits the installation of four optical cables in a conduit occupied by the 900-pair wire cable. Together with the reduced size goes an enormous reduction in weight: 3.6 kg of optical cable can replace 94.3 kg of copper cable. The size and weight reduction make optical cables ideal for communication links in aircraft and ships. Of course, there are also cost savings associated with lower weight and less bulk.

4. Immunity to inductive interference exists because the fiber is a dielectric, so changing electric or magnetic fields do not induce interfering signals. That is, the optical fiber does not act as an antenna to pick up radio-frequency interference (RFI), electromagnetic interference (EMI), or electromagnetic pulses (EMP). Consequently, cross-talk, which is the coupling of information from one pair of copper wires to another nearby, is virtually nonexistent in optical fiber cables. The result is noise-free transmission.

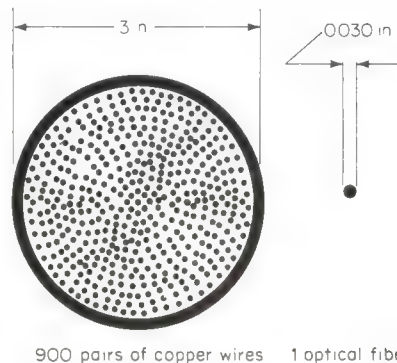
5. Excellent signal confinement occurs because the optical guide is designed to trap the optical electromagnetic wave within the fiber. Hence, external electric and magnetic fields are virtually nonexistent. Such a phenomenon is important for secure communication links required by military organizations and desired by cable television services.

To employ these advantages, the designer must assemble a communication link which satisfies the system requirements for analog signal-to-noise ratio (SNR) or digital bit error rate (BER), bandwidth or data rate, transmission distance, and environment. To accomplish this goal, a number of factors must be considered. The amount of optical power coupled into the fiber and the receiver noise sets a limit to the maximum received signal quality. Losses of optical power in the fiber degrade this performance quantity. The longer the fiber, the worse these losses become.

Important to the bandwidth requirement are the contributions to the risetime and pulse broadening made by the system. The risetime  $T_r$  is the time required for a pulse to rise from

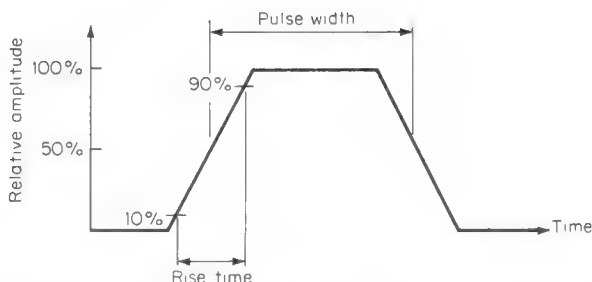


**Fig. 21.3** Comparison of attenuation as a function of modulation frequency for several coaxial cables and an optical fiber. (Courtesy AMP, Inc., Harrisburg, Pa.)



**Fig. 21.4** Size comparison between a 900-pair copper telephone cable and an optical cable with the same information capacity.





**Fig. 21.5** The meaning of risetime and pulse width of a pulse waveform.

10 to 90 percent of its maximum value, as indicated in Fig. 21.5. The pulse broadening  $T_{1/2}$  refers to increases in the pulse width measured between the one-half maximum value. For a system with a transmitter, fiber, and receiver in cascade, the composite risetime is given by

$$T_r^2(\text{sys}) = (1.1)^2 [T_r^2(\text{trans}) + T_r^2(\text{fiber}) + T_r^2(\text{rec})] \quad (21.2)$$

The pulse broadening for the system is determined in a similar manner. The corresponding equivalent bandwidth is approximately

$$B = \frac{0.32}{T_r} \quad (21.3a)$$

or

$$B = \frac{0.44}{T_{1/2}} \quad (21.3b)$$

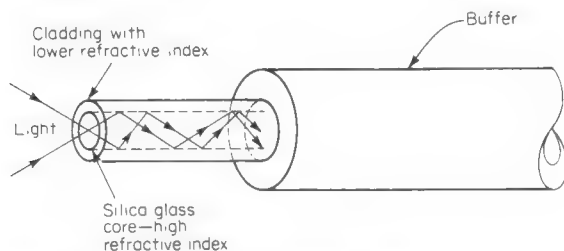
in terms of pulse broadening. These equations are used frequently in optical system design.

## 21.2 OPTICAL FIBERS AND CABLES

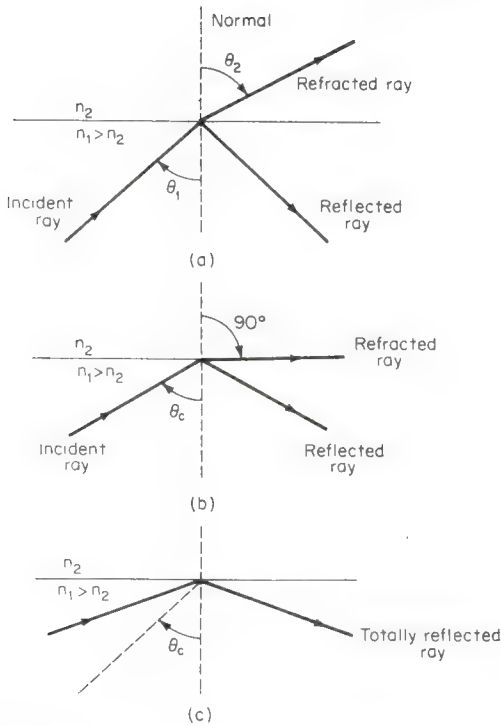
Driving the development of fiber-optic systems has been the rapid advance in low-loss, wide-bandwidth optical fibers. These thin glass or plastic strands, with diameters between 0.003 and 0.5 mm, conduct the modulated optical carrier from the transmitter to the detector. The choice of the proper optical cable to accomplish this task depends on the fiber light-guiding ability, attenuation, bandwidth, cross-sectional structure (cabled fibers), and cable configuration (fiber connectors).

**LIGHT GUIDING** The simplest fiber type consists of a cylindrical core of glass (silica) or plastic with index of refraction  $n_1$  surrounded by a concentric layer of glass or plastic of index  $n_2$ , as shown in Fig. 21.6. The index of refraction  $n$  is defined as

$$n = \frac{\text{speed of light in vacuum}}{\text{speed of light in material}} \quad (21.4)$$



**Fig. 21.6** Basic structure of a step-index fiber.



**Fig. 21.7** Reflection and refraction of light rays at a boundary between two different materials for (a) light incident at an angle less than the critical angle, (b) light incident at the critical angle, and (c) light incident at an angle greater than the critical angle.

Its value ranges from 1.45 to 1.55 for silica glass. Hence, light travels approximately 33 percent slower in silica than in air. The second layer is called the *cladding*, and it has a refractive index that is less than that of the core by a few percent. Such step-index structures are identified as either all-plastic, plastic-clad silica (PCS) or glass-clad silica (CGS). Core cladding sizes range from 30 to 125  $\mu\text{m}$  and 400 to 600  $\mu\text{m}$ , respectively. Additional layers, called buffers, are added to protect the cladding and to strengthen the fiber.

Optical power coupled into the fiber core is prevented from escaping the core by being totally reflected at the core boundary. Therefore, the light is guided along the fiber core. The basis for light guiding is illustrated in Fig. 21.7. Notice in Fig. 21.7a the light ray directed toward a boundary at an angle  $\theta_1$ . Some of the light is reflected at the same angle  $\theta_1$ . The rest of the light crosses the boundary into the second medium as a refracted ray which travels at an angle  $\theta_2$ , given by Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (21.5)$$

For this illustration  $n_2 > n_1$ , so that  $\theta_2 > \theta_1$ .

**example 21.2** If the core index is  $n_1 = 1.5$ , the cladding index  $n_2 = 1.4$ , and the angle of incidence  $\theta_1 = 20^\circ$ , what is the angle of refraction?

**solution** By substituting the given values into Eq. (21.5), we get

$$1.5 \sin 20^\circ = 1.4 \sin \theta_2$$

$$\theta_2 = 21.5^\circ$$

and

Note that only part of the light is reflected at the boundary; the rest is lost owing to refraction. The objective is to eliminate the loss of light power owing to refraction.

## 21-6 Fiber-Optic Systems

For light rays that are directed into a material with a lower refractive index, there is a certain angle of incidence for which the refracted angle is  $90^\circ$ . This specific angle of incidence is called the *critical angle*  $\theta_c$  and is illustrated in Fig. 21.6b.

**example 21.3** For the index values given in Example 21.2, determine the critical angle.

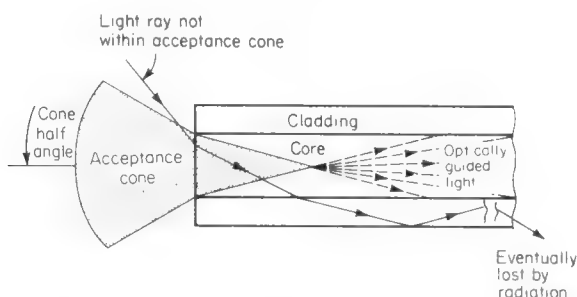
**solution**

$$1.5 \sin \theta_1 = 1.4 \sin 90^\circ$$

and

$$\theta_c = 69^\circ$$

At angles of incidence greater than  $\theta_c$ , total internal reflection of light occurs at the core-cladding boundary in a fiber, as shown in Fig. 21.7c.



**Fig. 21.8** Acceptance cone for a step-index fiber. Light that enters the fiber core at an angle greater than the cone half-angle is lost through refraction at the core-cladding boundary.

Therefore, light rays admitted to the fiber core only within a range of angles are guided by the fiber. This range defines an acceptance cone which is illustrated in Fig. 21.8. This cone is specified by the numerical aperture (N.A.) given by

$$\text{N.A.} = \sin \theta = \sqrt{n_1^2 - n_2^2} \quad (21.6)$$

where  $\theta$  is the cone half-angle. For most fibers, the cladding index is only a few percent smaller than for the core, so Eq. (21.6) may be approximated as

$$\begin{aligned} \text{N.A.} &\approx \sqrt{2n_1(n_1 - n_2)} \\ &\approx \sqrt{2n_1 \Delta} \end{aligned} \quad (21.7)$$

where  $\Delta = n_1 - n_2$ . Step-index fibers have N.A. values between 0.20 and 0.60.

**example 21.4** If the index difference is  $\Delta = 0.05$  and  $n_1 = 1.5$ , find the numerical aperture.

**solution**

$$\begin{aligned} \text{N.A.} &= 2(1.5)(0.05) \\ &= 0.387 \end{aligned}$$

This corresponds to an acceptance cone angle of  $45.6^\circ$ .

An important aspect of the numerical aperture is that it describes the light-capturing power of the fiber core. The larger the numerical aperture, the greater the optical power coupled into the fiber core, given approximately by

$$P_c = \frac{m + 1}{2} \frac{A_f}{A_e} (\text{N.A.})^2 P_o \quad (21.8)$$

where  $m$  = transmitter optical spatial distribution factor (discussed later)

$A_f$  = fiber core area

$A_e$  = light source emitting area

$P_o$  = total transmitter optical output power

If  $A_f > A_e$ , then  $A_f/A_e = 1$ .

**example 21.5** For an LED with  $P_o = 5 \text{ mW}$ ,  $m = 2$ , and  $A_e = 50 \times 150 \mu\text{m}^2$  coupled to a fiber with a  $250\text{-}\mu\text{m}$ -diameter core and  $\text{N.A.} = 0.45$ , determine the coupled power coupled into the fiber.

**solution**

$$\text{So } \frac{A_f}{A_e} = \frac{\pi(125 \mu\text{m})^2}{50 \times 150 \mu\text{m}^2} = 6.5 > 1$$

$$P_c = \frac{3}{2} (1)(0.45)^2 (5 \text{ mW}) = 1.52 \text{ mW}$$

The coupling loss in decibels is given by

$$\begin{aligned} \text{Loss} &= 10 \log \frac{5 \text{ mW}}{1.01 \text{ mW}} \\ &= 5.17 \text{ dB} \end{aligned}$$

**example 21.6** If the core diameter in Example 21.5 is reduced to  $50 \mu\text{m}$  and the  $\text{N.A.}$  to 0.2, which is typical for telecommunications fibers, what is the coupled power?

**solution**

$$\frac{A_f}{A_e} = \frac{\pi(25 \mu\text{m})^2}{50 \times 150 \mu\text{m}^2} = 0.26$$

$$\text{then } P_c = \frac{2+1}{2} (0.26)(0.20)^2 (5 \text{ mW}) = 80 \mu\text{W}$$

The coupling loss is 18 dB.

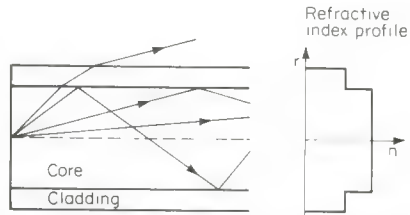
These examples illustrate the idea that as the  $\text{N.A.}$  decreases, the losses increase.

A problem with the step-index structure is that light rays may travel along several different paths, as shown in Fig. 21.9. For this reason, the step-index fiber is called a *multimode guide* and suffers from multimode dispersion. Since light energy traveling along two different paths propagates with different times, an input pulse of light is broadened, as shown in Fig. 21.10. Therefore, there is a limit to the pulse or data rate and to the fiber bandwidth. At present in commercially available step-index fibers, the bandwidth is limited to approximately 25 MHz for a 1-km length. This property is examined in more detail later.

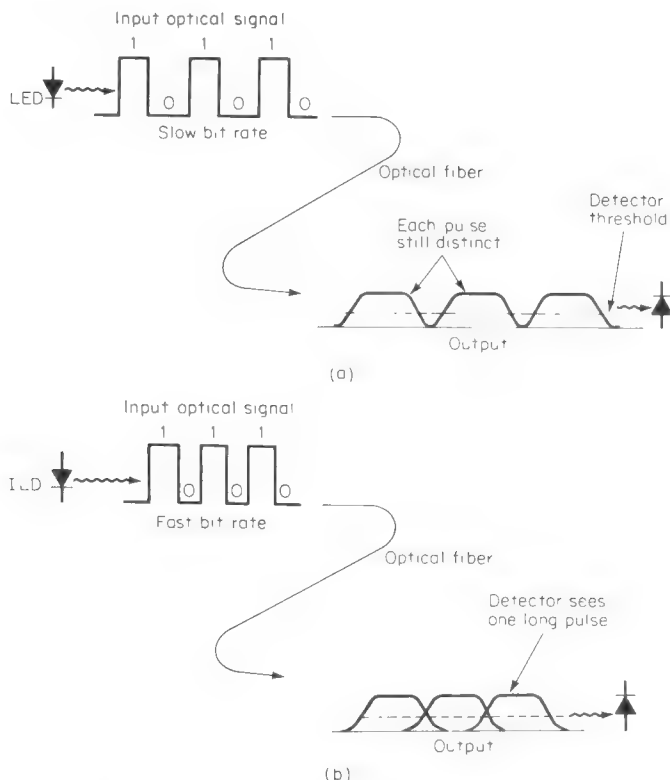
To increase the bandwidth or pulse rate, the fiber core can be manufactured with the index of refraction gradually reduced as the core radius increases. The result is a graded-index fiber shown in Fig. 21.11. The effect is that light which travels farther from the core travels faster than that along the core center. For an appropriate index profile, the propagation time for the various modes is nearly equalized to yield a greatly reduced multimode dispersion term. Optical bandwidths for graded-index fibers range from 200 to over 1500 MHz for 1 km of fiber. A core buffer diameter of 50 to  $125 \mu\text{m}$  is becoming a standard for telecommunications applications. Numerical aperture values range from 0.17 to 0.28.

If the core diameter is less than approximately 3.2 times the wavelength of the light signal in a step-index guide, then the light travels along one path, that is, in a single mode. This offers the potential for modulation bandwidths exceeding 1000 GHz. Single-mode fibers have a core diameter of 4 to  $8 \mu\text{m}$  for 850-nm light. Because the core diameter of such fiber is so small, efficient light coupling between it and the optical transmitter or other fibers is difficult. For this reason, single-mode fibers are used primarily in laboratory demonstration applications. Single-mode fibers have cladding diameters in the range of 100 to  $125 \mu\text{m}$  and numerical apertures of about 0.1.

**ATTENUATION** Even with the best fibers available some of the light is lost by internal absorption and scattering, which attenuates the optical power. The optical intensity decreases geometrically as the signal travels from the source to the detector. For example,



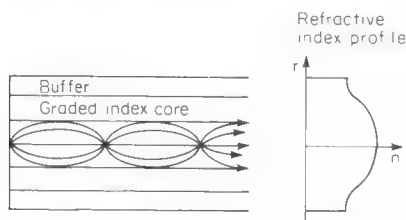
**Fig. 21.9** Propagation paths for light in a step-index fiber.



**Fig. 21.10** Effect of multimode dispersion on digital pulse waveforms for (a) a low-dispersion fiber and (b) a high-dispersion fiber. (Courtesy AMP, Inc., Harrisburg, Pa.)

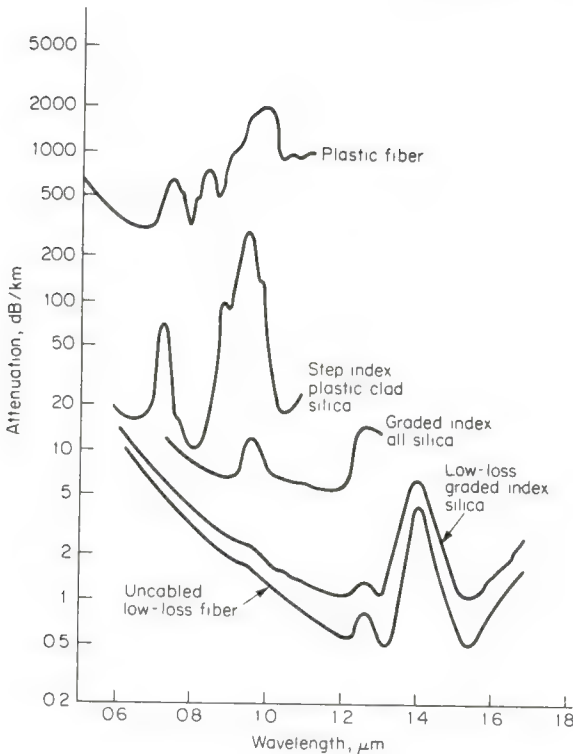
if the strength of the light falls to half its original value after the light has traveled 1 km, the intensity will fall to one-fourth of the original value at the end of the second kilometer, and so on. For long-distance transmission, the light source should be as powerful as possible and the cable losses or attenuation as low as possible, other things being equal.

Attenuation, or loss of light with distance, is expressed in decibels per kilometer (dB/km). Values range from less than 1 dB/km for low-loss telecommunications cables to 1000 dB/km for plastic fibers useful for short-distance transmission. Typical curves of attenuation versus optical wavelength for several fiber types are shown in Fig. 21.12. The attenuation is caused by four principal mechanisms: material absorption, material scattering, waveguide scattering, and leaky modes. The differences in the various curves represent the degree to which these imperfections have been reduced. For long fiber lengths, it is important to use an LED or laser diode which emits light with a wavelength where the fiber attenuation is a minimum. Properties of several fiber types are shown in Table 21.1.



**Fig. 21.11** Propagation paths for light in a graded-index fiber.

Material absorption is caused by impurities in the core of the fiber which are exceedingly difficult to remove. The main impurities are the ions  $\text{OH}^-$ ,  $\text{Fe}^{2+}$ , and  $\text{Cr}^{3+}$ , which in concentrations by weight of 1 part in  $10^9$  to 1.25 parts in  $10^6$  result in a peak loss of 1 dB/km. The absorption peaks near 950 and 1400 nm for silica-based fibers are due to the  $\text{OH}^-$  ions. Actual fibers con-



**Fig. 21.12** Dependence of fiber attenuation on the optical wavelength for several fiber types.

tain varying amounts of these impurities depending on glass purification processes. Material or Rayleigh scattering is caused by frozen-in fluctuations in glass composition as well as glass-phase separations, which are small relative to the wavelength of light. This scattering varies as  $1/\lambda^4$  and explains the decrease in attenuation with increasing wavelength for low-loss fibers.

Waveguide scattering results from geometric variations in the size of the core along its length, which in turn cause transfer of energy from one guided mode to another or from a guided mode to one that leaks into the cladding. In the former case, each mode may have different absorption and radiation coefficients. For a poorly made fiber, waveguide scattering losses are 10 dB/km.

The solution of Maxwell's equations which describe light propagation shows that even though light is totally reflected at the core boundary, the optical electromagnetic field extends into the cladding. (See Fig. 21.13.) This field is referred to as the *evanescent* field because its amplitude vanishes exponentially with radial distance from the core center. Attenuation or loss of the evanescent field owing to radiation or absorption from the cladding causes attenuation of the guided light in the core. Thus, the cladding thickness must be great enough that the evanescent field at the outer cladding surface is insignificant. Also, the cladding material should be almost as pure as that for the core. One of the reasons that plastic clad fibers have higher attenuation than all-glass fibers is that the plastic cladding is more lossy than glass cladding.

The importance of fiber attenuation in a fiber-optic link can be illustrated graphically. The optical power present in a fiber at a given distance  $L$  from the source can be computed as

$$P_f = P_c 10^{-\alpha L} \quad (21.9)$$

or as

$$P_f \text{ (dB}_m\text{)} = P_c \text{ (dB}_m\text{)} - \alpha L \quad (21.10)$$



**TABLE 21.1 Properties of Selected Optical Fibers**

Attenuation (dB/km @ $\lambda$ in nm)	Bandwidth (MHz · km)	Numerical aperture	Core/ clad diameter ( $\mu$ m)	Maximum length (km)	Manufacturer	Model no.
Single mode (step index)						
0.7 @ 1300	10 <sup>4</sup>		10/125	3	Fujikura Cable	SML 10/125-7
2.5 @ 850	10 <sup>4</sup>		6/125	3	Fujikura Cable	SM 6/125-25
5 @ 850	>1000	0.11	4/80	3	ITT, Electro- Optical	T 1601
3-20	≥80 000	0.12	5/100	1	Galite	7000
Graded index						
1 @ 1200	500	0.2	50/125	3	Valtec	ESP
1/2.5 @	400-1000	0.2	50/125	1.1	Corning Glass Works	2504D2510D
2/3.5 @	200-1000	0.2	50/125	1.1	Corning Glass Works	3502D-3510D
2-10	100-1000	≥0.21	63/125	3	Galite	6000 series
2.5 @ 850	200-800	0.2	50/125	3	Valtec	
3 @ 850	600-1500	0.21	50/125	1.1	Corning Glass Works	3006F-3015F
3 @ 850	800	0.2	50/125	5-10	Sumitomo Electric	EG-5/4060
3 @ 850	800	0.17	50/125	1	Dainichi-Nippon Cables	SM50G-SY-3L
3.5 @ 850	400-800	0.21	50/125	1.1	Corning Glass Works	3504F-3508F
3.5 @ 850	500	0.2	50/125	5-10	Sumitomo Electric	EG-5/4040
3.5 @ 850	>1000	0.22	50/125	10	ITT, Electro- Optical	T 1211
4 @ 850	200	0.2	50/125	3	Valtec	420
4-5 @ 840	300	0.18	50/125	1.5	Fort	LG
4 @ 840	500	0.17	60/125	2	Thomson CSF, LTT	GL 4
4 @ 850	800	0.21	50/125	1.1	Corning Glass Works	4008F
4 @ 850	300	0.2	50/125	3	Optronics	GIT
4 @ 850	>1000	0.22	50/125	10	ITT, Electro- Optical	T 1209
5 @ 850	400	0.21	50/125	1.1	Corning Glass Works	5004F
5-8 @ 820	200-400	0.16	50/125	1	Times Fiber Commun	G-5000
5 @ 850	400	0.25	62/125	1.0	General Fiber Optics	G-125/62-5
5-8 @ 850	200-800	0.22	50/125	10	ITT, Electro- Optical	T 1200 series
5 @ 850	800	0.17	60/150	1	Dainichi-Nippon Cables	SM60G-SY-5L
6-8	150	0.22	50/125	2	Standard Telephone	MA 3211-MA 2211
6 @ 800	200	0.18	65/130	5	Fiberoptic Cable	G1-2/1-6UJ
6 @ 850	200-400	0.21	50/125	1.1	Corning Glass Works	6002F-6004F
6 @ 900	200	0.2	50/125	10	Fibronics	FGI-50-D
<7 @ 820	400	0.21	50/125	1.1	Brand-Rex	Type II
Step index						
2-10	10-20	≥0.21	63/125	3	Galite	5000 series
3-10	10-15	≥0.3	125/200	2	Galite	5020 series
4 @ 820	30	0.21	90/125	20	Quartz & Silice	QSF 125AS-S
4 @ 820	50	0.25	250/260	2	Fiberoptic Cable	Q1-LO-100J

TABLE 21.1 Properties of Selected Optical Fibers (Continued)

Attenuation (dB/km@ $\lambda$ in nm)	Bandwidth (MHz·km)	Numerical aperture	Core/ clad diameter ( $\mu$ m)	Maximum length (km)	Manufacturer	Model no.
Step index						
5@820	30	0.4	125/300	25	Quartz & Silice	QSF 125A
5-20@820	25	0.3	200/375	1	Times Fiber Commun	P-2000
5@850	>30	0.25	50/125	10	ITT, Electro- Optical	T 1101
5@850	30	0.17	100/150	1	Dainichi-Nippon Cables	SM 100-SY-5L
5@850	30	0.19	200/400	3	Optronics	PC 200A
6@820	40	0.25	125/150	5	Fiberoptic Cable	Q1-LO-5UJ
<7@820	10	0.22	600	1.1	Brand-Rex	Type X
7@700	20	0.3	100/140	1.0	General Fiber Optics	S-140/100-7
10@850	15	0.30	100/140	1	ITT Components Group	SG5
10@790	20	0.3	200/400	2	BOTec	EIS-200
10@800	25	0.3	250/550	3	Valtec	PC 10
10@820	25	0.24	200/300	4	Focom Systems	200/PCSI/10
10@850	60	0.2	50/125	3	Optronics	SLT
12@800	20	0.3	250/430	3.5	Fibronics	FSI-250
<15@900	>10	0.25	100/150	1.5	Pilkington	15F150
15	5-15	$\geq 0.4$	204/305	2	Galite	4000LC
20@790	25	0.23	200/400	3	Fiberguide Industries	Superguide-3
20@820	25	0.25	250/260	2	Fiberoptic Cable	Q1-MD-10UJ
25@820	22	0.24	250/380	2	Focom Systems	200/PCSI/25
25@900	30	0.25	113/147	1	ITT Components Group	SG4
30@820	35	0.4	125/300	25	Quartz & Silice	QSF 125C
30@820		0.42	200/600		DuPont	S-120 type 30
<40@900	6	0.48	200/250	10	Pilkington	40F250
40@820	15	0.24	400/550	1.2	Focom Systems	400/PCSI/40
45@670	20	0.25	250/310	1	ITT Components Group	PSI
<50@820	25	0.22	200/400	1.1	Brand-Rex	Type VII
50@660	20	0.23	200/400	3	Fiberguide Industries	Practiguide
50@790	20	0.3	700/400	2	BOTec	EIS-200H
50@820	20	0.4	400/550	5	Quartz & Silice	QSF 400B

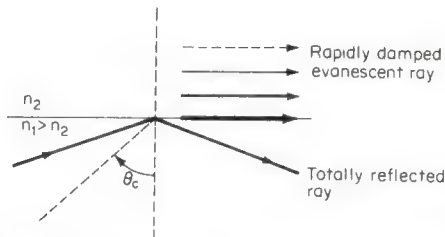
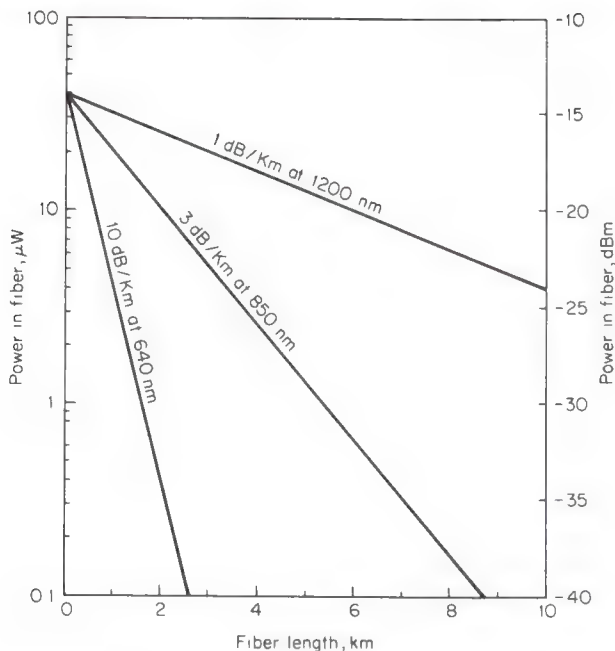


Fig. 21.13 Propagation of the evanescent light field for light which is totally reflected at a boundary between two different materials. None of the evanescent light is lost by radiation if the medium in which it travels extends to infinity.



**Fig. 21.14** The effect of different fiber attenuation values or different source wavelength values on the decrease of optical power in a fiber with increasing fiber length. The initial power level is  $40 \mu\text{W}$ . For a receiver sensitivity of  $0.7 \mu\text{W}$  (dashed horizontal line), the maximum fiber length can be estimated.

where  $\alpha$  is the fiber attenuation in decibels per kilometer and  $\text{dB}_m$  is a unit for power referenced to  $1 \text{ mW}$ :

$$\text{dB}_m = 10 \log \frac{\text{power}}{1 \text{ mW}}$$

Note that Eq. (21.10) represents a straight line on a semilog graph of  $P_f$  versus  $L$ . Several graphs are shown in Fig. 21.14 for different fiber attenuation values. Here, the power coupled into the fiber is taken to be  $40 \mu\text{W}$  ( $-14 \text{ dB}_m$ ). The three lines drawn on the graph indicate the optical power as a function of length of silica fiber for attenuation values at three different optical wavelengths. If the received optical power must exceed  $0.7 \mu\text{W}$  ( $-31.5 \text{ dB}_m$ ), then the maximum fiber lengths possible are

1.8 km for a 640-nm-wavelength source

5.9 km for a 850-nm-wavelength source

17.5 km (off scale) for a 1200-nm-wavelength source

The importance of matching the transmitter optical wavelength to the wavelength for minimum fiber attenuation should be apparent.

**FIBER BANDWIDTH** In theory, an optical carrier provides a potential modulation bandwidth exceeding  $10 \text{ THz}$ . Practical optical guides, however, only permit the transmission of information at much lower rates, because of pulse-broadening effects. The bandwidth degradation may be described in terms of a pulse-broadening factor  $\tau_{1/2}$  with units of nanoseconds per kilometer ( $\text{ns/km}$ ), a risetime factor  $\tau_r$  with units of nanoseconds per kilometer, or in terms of a bandwidth-length product (BLP) with units of megahertz-kilometers ( $\text{MHz} \cdot \text{km}$ ). To a first approximation, the pulse-broadening and risetime terms vary linearly with fiber length while the bandwidth varies inversely. For example, step index fiber with

a bandwidth of 25 MHz at 1 km has a bandwidth of 6.25 MHz at 4-km length. These three terms are related by

$$\text{BLP} = \frac{0.32}{\tau_r} \quad (21.11)$$

$$\text{BLP} = \frac{0.44}{\tau_{1/2}} \quad (21.12)$$

Three effects contribute to this degradation of bandwidth: (1) multimode dispersion, (2) material dispersion, and (3) waveguide dispersion. Each is described below.

Recall from the discussion of fiber types that multimode dispersion may occur if the light travels along different paths in the fiber. For the step-index fiber the contribution to the pulse broadening varies as

$$\tau_{1/2} \propto (\text{NA})^2 \quad (21.13)$$

A large numerical aperture increases the optical power coupled into the fiber but reduces the bandwidth. The design of the step-index fiber represents a compromise between these two conflicting terms. For this reason the cladding index of refraction is only a few percent smaller than that for the core.

For graded-index fibers, the multimode dispersion is reduced significantly by equalizing the propagation time for each mode. However, this equalization depends critically on the index profile and on the transmitter wavelength. Although theory predicts that the pulse broadening for graded-index fibers may be 10 000 times smaller than that for step-index fibers, this performance level is not achieved in practice.

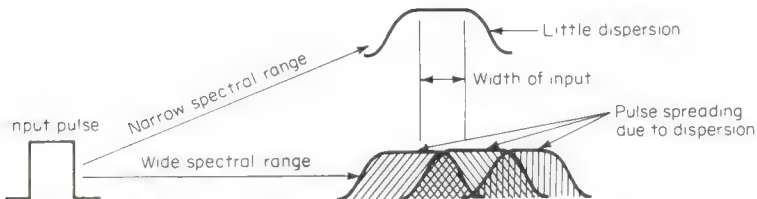
The second pulse-broadening effect is material dispersion which is important to wide-bandwidth applications. It arises from the wavelength dependence of the refractive index. Optical sources emit over a band of wavelengths—not just one. The result is that different optical frequencies in a pulse of light travel at varying relative speeds within the fiber. Hence, pulse broadening occurs as illustrated in Fig. 21.15. For LEDs and ILDs in the 800 to 900 nm range, the pulse broadening is approximately 5 and 0.5 ns/km, respectively. Hence, even if multimode dispersion were to be eliminated, material dispersion would limit the fiber bandwidth to 90 MHz·km with an LED transmitter and to 900 MHz·km for an ILD transmitter in this wavelength range.

Fortunately, fibers may be produced for which material dispersion is reduced to near zero near a wavelength of 1300 nm. Hence, there is much interest in graded-index fibers for applications between 1200 and 1300 nm where fiber attenuation is low also (refer to Fig. 21.12).

The remaining contribution to pulse broadening is the waveguide dispersion effect. Any carrier frequency  $f_c$  which is amplitude-modulated by a signal frequency  $f_s$  has two sideband frequencies produced: (1) the upper sideband at  $f_c + f_s$  and (2) the lower sideband at  $f_c - f_s$ . The upper sideband travels in the optical guide with a slightly different speed than the lower sideband. Again, the result is pulse broadening. However, because the modulation or signal frequencies are much smaller than the carrier frequency, this effect may be ignored for all applications not involving single-mode fibers.

The combined pulse broadening for these fiber dispersion terms is given by

$$\tau_{1/2}^2 (\text{fiber}) = \tau_{1/2}^2 (\text{modal}) + \tau_{1/2}^2 (\text{material}) + \tau_{1/2}^2 (\text{waveguide}) \quad (21.14)$$



**Fig. 21.15** Pulse broadening owing to material dispersion. The light energy at each wavelength may travel at a slightly different speed and arrive at the end of the fiber at a different time. (Courtesy AMP, Inc., Harrisburg, Pa.)

Generally, a system excited by an LED will have a greater pulse broadening and lower bandwidth owing to material dispersion than a system with an ILD source.

In general, caution should be exercised in system engineering with BLPs or pulse-broadening values from manufacturers' literature for the following reasons:

1. Many manufacturers measure the pulse-broadening effects with an ILD source which overestimates the fiber bandwidth for an LED-excited system.

2. The BLP is a function of fiber length. Many manufacturers measure its value with short links and extrapolate to 1 km.

3. The values obtained for  $\tau_r$  and the BLP depend critically on the type of optical source and the source-to-fiber coupling technique. It is common to find that different laboratories will obtain values that vary by a factor of 2 or more for the same fiber type and length as a result of differences in optical coupling schemes.

For these reasons, a fiber dispersion factor should specify the measurement conditions including the fiber length and the source coupling scheme. Bandwidth-length products for several fiber types are given in Table 21.1.

**example 21.7** An 820-nm LED is used as the source for a graded-index fiber which has a 200 MHz·km bandwidth-length product. Determine the effective BLP.

**solution** The effective BLP is less because the 5-ns/km material dispersion term must be included. First determine the modal dispersion value from Eq. (21.12):

$$\begin{aligned}\tau_{1/2} &= \frac{0.44}{200 \text{ MHz} \cdot \text{km}} \\ &= 2.2 \text{ ns/km}\end{aligned}$$

From Eq. (21.14) the effective pulse broadening becomes

$$\tau_{1/2}^2 = (2.2 \text{ ns/km})^2 + (5 \text{ ns/km})^2$$

and

$$\tau_{1/2} = 5.5 \text{ ns/km}$$

The bandwidth-length product becomes

$$\begin{aligned}\text{BLP} &= \frac{0.44}{5.5 \text{ ns/km}} \\ &= 80 \text{ MHz} \cdot \text{km}\end{aligned}$$

**CABLED FIBERS** A cladded, buffered fiber is not protected properly for the environments likely to be encountered in most applications. Provision must be made for the following:

1. Mechanical abuse from impact, crushing, flexure, and tension which arise from installation and service

2. Temperature extremes

3. Water ingress which may lead to stress corrosion or to stress owing to freezing

4. Chemical deterioration of cable materials

5. Radiation-induced attenuation

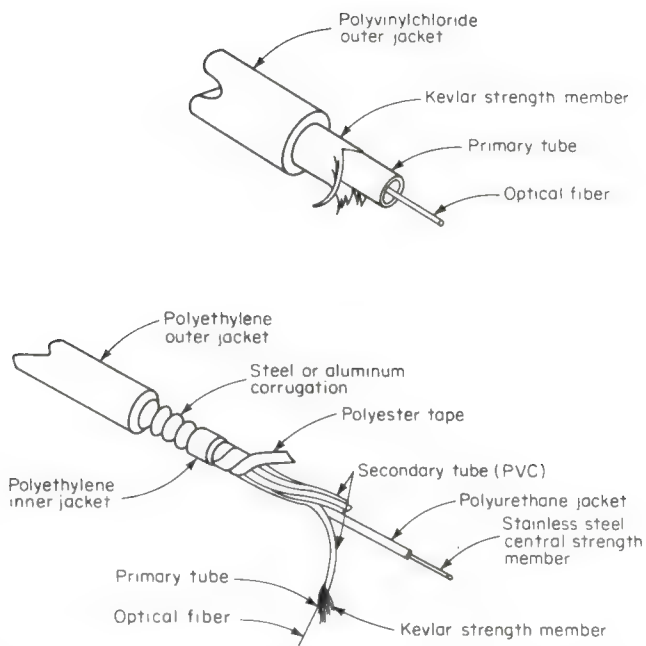
6. Damage from gnawing rodents

A variety of cable structures are available to accommodate or reduce the effect of these stresses. Examples are given in Figs. 21.16 and 21.17.

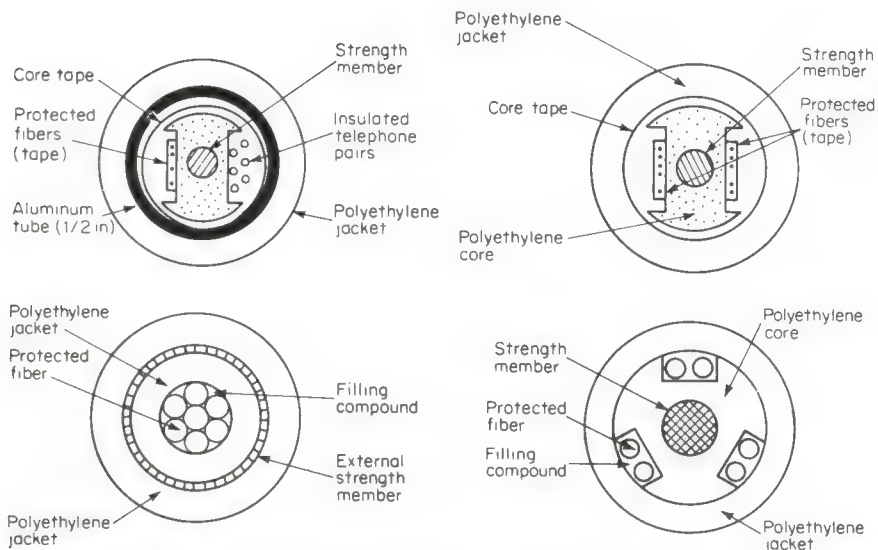
All cables have a maximum tensile load. Even though the tensile strength of individual fiber is in the range of  $2 \times 10^5$  to  $3 \times 10^5$  kg<sub>f</sub>/m<sup>2</sup>, the small cross-sectional area yields a maximum load of only 1.0 to 1.5 kg<sub>f</sub>, or 2 to 3 lb. Therefore, optical fiber must be combined with strands of Kevlar\*, fiber glass, or steel to provide the tensile strength required for installation. Cable structures for use with direct-bury, underground duct, plowed, or aerial installation techniques are available. Types of specific cable applications are listed in Table 21.2.

To prevent the introduction of permanent effects in the fiber properties during installation, the maximum tensile force should be limited to about 44 kg<sub>f</sub> (100 lb). This should be adequate for cable installation for most applications. For example, a tensile load of 13 kg<sub>f</sub> is normally required to pull a 1-km cable of 1.6-cm diameter through a multiply curved conduit. Some firms use a polyethylene tube or a lubricated metal duct tube to reduce friction in pulling cables through ducts.

\* Kevlar is a registered trademark of DuPont, Inc.



**Fig. 21.16** Structures for a single-fiber cable and for a multiple-fiber cable.



**Fig. 21.17** Cross sections of different structures for multiple-fiber cables. (Courtesy General Cable Company, Woodbridge, N.J.)



TABLE 21.2 Characteristics of Selected Fiber-Optic Cables

Attenuation (dB/km @ $\lambda$ in nm)	Bandwidth (MHz·km)	Number of fibers	Tensile strength (kgf)	Maximum length (km)	Applications and features	Manufacturer	Typical product
0.7@1300	10	1-48	50-300	2	Interurban trunks	Fujikura Cable	NA/(5-60)
3-10@820	200-1500	5-60	300	1.1	General telecommunications	Compagnie Lyonnaise	SM50/6-LAP
3@850	30	6	150	1	All nonmetallic	Dainichi-Nippon	2-804350NM
3.5-8	200-800	4-12	400	2	Aerial	General Cable	T 2501
3.5-10@850	30-1000	1-19	400	4	Duct	ITT	T 2500
3.5-10@850	30-1000	1-19	400	4	Buried	ITT	T 2502
3.5-10@850	30-1000	1-19	400	1	Simplex	ITT	T 3000-3101
3.5-10@850	30-1000	1	30	4	Duplex	ITT	T 3001-3102
3.5-10@850	30-1000	2	60	4	General purpose	ITT	T 3500
3.5-10@850	30-1000	1-7	100	4	High strength, lightweight	Valtec	MP-MG05
5@850	200	1-6	100	3	Military	Valtec	FC-MG05
5@850	200	6	50	3	Heavy-duty design	Valtec	HD-MG05
5@850	200	2	50	3	Low cost, intrabuilding	Valtec	GP-MG05-01
5@850	200	1	25	3	Telephone trunking	Valtec	TC-MG05
5@850	200	4	225-350	3	General purpose	Siecor	622
6@820	400	1-10	300-350	1	Oval, reinforced	Times Fiber	IN-6
6@820	400@820	1-8	2-20	1	Oval, reinforced air and duct	Times Fiber	AD-6
6@820	400@820	2-12	2-45	1	Communications & instrumentation	Belden	227002
8@850	200	2	180-320	1	Communications & instrumentation	Belden	226018
9	20	2-18	113-317	1	Data communications	Siecor	242
10@820	10	2	400	1	General purpose	Siecor	133
10@820	100	1	30	1	Oval, reinforced	Times Fiber	IN-10
10@820	200	1-8	3-20	1	Communications & instrumentation	Belden	220001
10@850	25	1-18	113-317	1	Communications & instrumentation	Belden	227001
10@850	200	1	110	3	Low cost, intrabuilding	Valtec	PC10
12@800	25	1-2	50	3	Small size, data communications	Valtec	DC-PC08
15@800	25	2	10	3	Flame-retardant	Focom Systems	200/PCS8R/25
25@820	25	1-8	100	2	Resists radiation	DuPont, Pifax	S-120
30@790	5	1-2	65-130	1.0	Flame-resistant	Siecor	155
35@850		1	150	1			

Cable bending can greatly affect fiber longevity. If a fiber with a cladding radius  $r$  is bent along a circumference of radius  $R$ , the outermost edge will be strained. The percentage strain is given by

$$\sigma_s = \left( \frac{R + 2r}{R + r} - 1 \right) \times 100\% \quad (21.15)$$

**example 21.8** A fiber with a cladding radius of  $125 \mu\text{m}$  is bent along a curve of radius  $R = 2 \text{ cm}$ . Calculate the corresponding strain.

**solution**

$$\begin{aligned} \sigma_s &= \left( \frac{2 \text{ cm} + 2 \times 0.0125 \text{ cm}}{2 \text{ cm} + 0.0125 \text{ cm}} - 1 \right) \times 100\% \\ &= 0.62\% \end{aligned}$$

To ensure fiber longevity the differential strain should not exceed 0.2 percent. For the fiber of Example 21.8, the bending radius  $R$  should be greater than 6.5 cm. When a company gives a minimum bending radius, it usually means the minimum radius that can be applied for a short time before the fiber breaks. Not given is the significant effect on fiber lifetime if the fiber is left in a curved condition with a radius greater than the minimum.

For protection against crushing or pests, the cables can be supplied with steel or aluminum armor, as in Fig. 21.16. However, the cable weight is increased greatly, and isolation from lightning is eliminated. Cables can be waterproofed, also, by the use of a petroleum-based jelly or a dry powder to fill cable voids.

Temperature extremes have an adverse effect on fiber attenuation. Plastic-clad and buffered fiber, on one hand, are usable to  $-30^\circ\text{C}$  where thermal expansion of the plastic causes stress that increases the attenuation markedly. Glass-clad fibers, on the other hand, are usable to below  $-50^\circ\text{C}$ . At high temperatures, the main problem is the durability of the cable jacketing materials: Kevlar, fiber glass, polyethylene, or polyurethane. Cables that meet the IEEE-383 flame-retardance specification can be obtained on special order. In general, the choice of cable materials and structure is dictated by the specific temperature range of the application. The manufacturer should be consulted for cable viability in any application environment.

**FIBER CONNECTING** Fiber-optic systems contain a number of individual components, source fibers, detectors, etc., which must be linked so that optical losses are minimal. There are three types of connecting methods:

1. The splice joins two fibers or two fiber bundles.
2. The connector links one fiber to another, to repeaters, or to end devices and is demountable from the rest of the fiber-optic transmission system.
3. The coupler distributes optical power received from an input port to several output ports.

The key problem in any single-fiber splice or connector is, first, to align the central axis of one fiber with that of another fiber and, second, to minimize separation of the fiber ends. For example, a  $76\text{-}\mu\text{m}$  (3-mil) misalignment of the axes for two  $380\text{-}\mu\text{m}$  fiber cores leads to a 2-dB loss; a  $76\text{-}\mu\text{m}$  end separation for the same fibers yields an additional 0.3-dB loss. The alignment should be done while fibers are accepted and joined with variations owing to manufacturing tolerances. For fibers with small core diameters (less than  $50 \mu\text{m}$ ), this process becomes critical.

**Splices** Several splicing techniques are illustrated for single fibers in Fig. 21.18 and for multiple fibers in Fig. 21.19. Of these, fusion or fiber welding with an electric arc has become widely implemented. At one time it was believed that fusion techniques were impractical for use in a nonlaboratory setting, particularly without highly trained workers. However, the recent success of cable splicers employed by United Telephone of Pennsylvania who had only 1 week of training and practice has dispelled this myth. Splices were made inside, outside, in breezy conditions, above ground, and in manholes.

In the fusion technique, the splicer works with the benefit of a 100X microscope, so the fibers appear about the size of pencils. The "weld" is made by melting the glass between electrodes. The key to a successful splice is to recognize the proper glass temperature by the color of the heated glass. As the glass fiber ends are heated, the color changes from orange

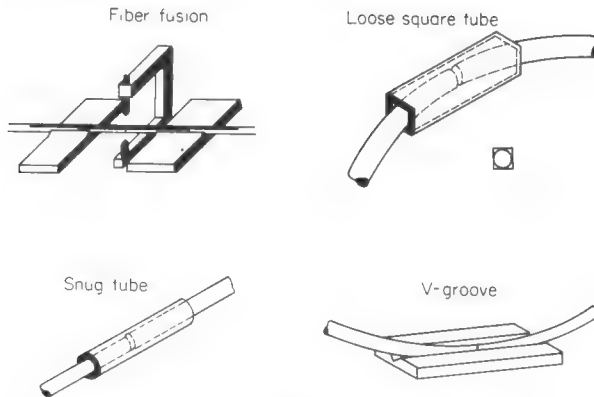


Fig. 21.18 Methods for splicing single optical fibers.

to pink to brilliant white. At that point the fibers are placed in contact for fusion. The average attenuation per splice ranges from 0.15 dB with a standard deviation of twice the mean value.

Another technique in Fig. 21.18 is the V groove which is preferentially etched into a silicon slice. A precision form is produced into which to lay the fibers. It also forms the basis for connecting multiple-fiber ribbons manufactured by Western Electric for the Bell system, as in Fig. 21.19a. Another, the precision sleeve splice, requires high degrees of accuracy and reproducibility of the inner sleeve dimensions and on the outer fiber diameter. For splices using index matching fluids and adhesives, precision tube splices have consistently achieved insertion losses less than 0.3 dB. The loose-tube splice uses a glass tube with a square cross section where the inner tube dimensions are purposely made larger than the outer fiber diameter. When the two fibers are bowed slightly, the loose tube rotates and holds the ends in place until the adhesive has cured. Various modifications of the loose-tube splice have been made with losses ranging from 0.3 to 0.5 dB.

**Connectors** If repeated connections to a fiber are required, a demountable connector is used. An example, produced by AMP, Inc., is shown in Fig. 21.20. The key element is the ferrule which is made of resilient plastic. On being mated, a bushing compresses the ferrule to produce alignment of the fibers with the connector axis. Use of connectors of this type requires that the cable jacket materials be stripped to expose the fiber. Epoxy is used to fix the fiber in the ferrule and the jacket to the ferrule body. After the epoxy is cured, the exposed fiber end is lapped with polishing papers to a mirror finish. Other connecting techniques, illustrated in Fig. 21.21, employ various alignment schemes to minimize losses. Regardless of the precision of connector components, fibers have minor geometric variations which add to potential misalignments even for perfectly mated male-female connector pieces. Commercially produced connectors have insertion losses ranging from 1.0 to 2.5 dB with costs ranging from \$3 for large fibers (diameters greater than 240  $\mu\text{m}$ ) to \$30 for precision single-fiber connectors (fiber diameter less than 140  $\mu\text{m}$ ).

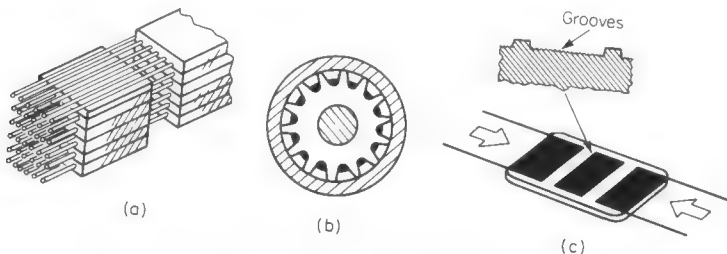
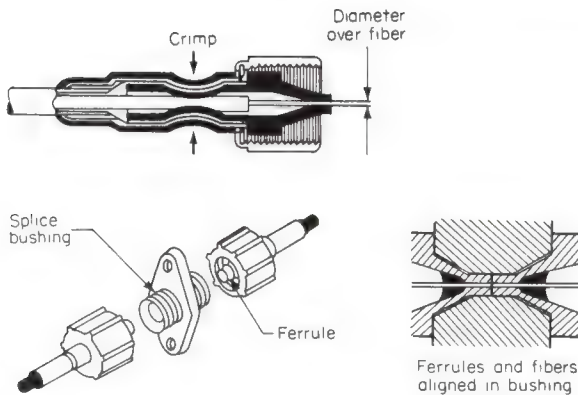
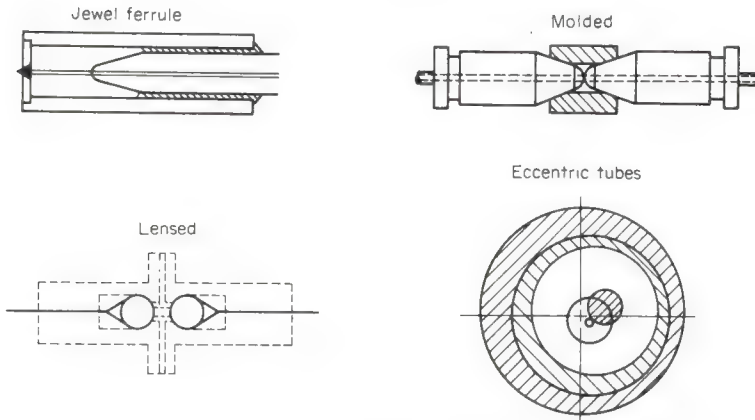


Fig. 21.19 Structures for splicing multiple optical fibers.



**Fig. 21.20** Example of a demountable connector. Fiber alignment is produced by compression of the resilient ferrule by the bushing. (Courtesy AMP, Inc., Harrisburg, Pa.)



**Fig. 21.21** Structures for connecting single optical fibers to minimize fiber alignment errors. The purpose is to maximize the coupling of light from one fiber into the other.

**Couplers** Couplers permit the distribution or combination of optical signals for any number of optical fiber lines or sources. Applications include data distribution networks, duplex communications with a single-fiber guide, and wavelength division multiplexing of several optical signals onto a single optical guide.

One common technique for producing a coupler is achieved by twisting lengths of unjacketed fiber together in a bundle. While the fiber bundle is being fused, the fiber ends are pulled to form a biconical taper. Typically, the fused region is bonded to a substrate and encapsulated in a protective package, as illustrated in Fig. 21.22.



**Fig. 21.22** Example of a biconical taper coupler.

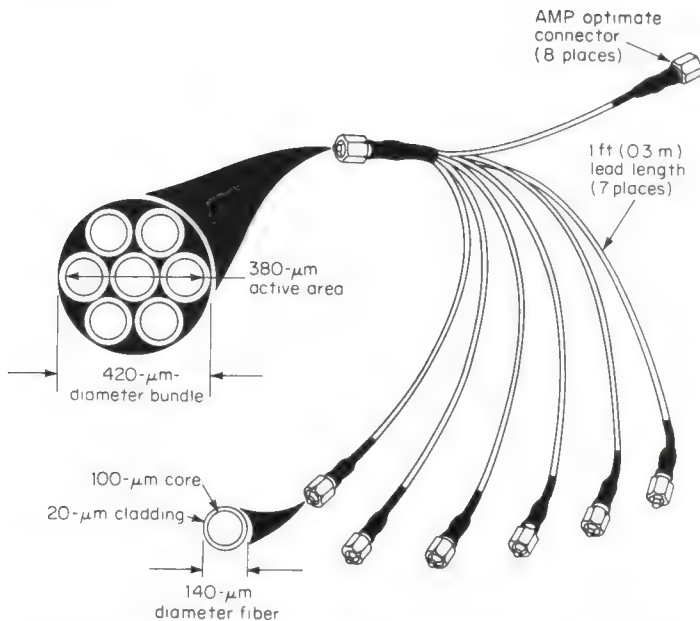


Fig. 21.23 A seven-way combiner-splitter. (Courtesy AMP, Inc., Harrisburg, Pa.)

A simple passive combiner or splitter can be configured with individual fibers, as shown in Fig. 21.23. Here seven fibers with a 140- $\mu\text{m}$  cladding diameter are mounted in a standard fiber-optic connector to form a 400- $\mu\text{m}$ -diameter bundle. The matching transmitter or receiver fiber should be a single fiber with a 400- $\mu\text{m}$  cladding diameter. Two such units coupled back to back with a length of 400- $\mu\text{m}$  fiber form a 14-arm star coupler.

Coupler technology is still in its infancy. Advances can be expected at a rapid pace which will increase the performance and reduce the cost of individual units.

**Connecting mismatch losses** Even if perfect fiber alignment and minimum core separation could be achieved with connecting techniques, additional losses may arise from mismatches between the core ends. For example, changes in the core refractive index at the end boundaries lead to light reflection. This attenuation, called *Fresnel loss*, is given by

$$\alpha_{\text{Fresnel}} = -10 \log \frac{4n_1n_2}{n_1 + n_2} \quad (21.16)$$

where  $n_1$  is the core refractive index and  $n_2$  is that for the medium outside the core end. For each silica-air boundary where  $n_1 = 1.5$  and  $n_2 = 1.00$ , the loss is

$$\alpha_{\text{Fresnel}} = 10 \log \frac{(n_1 + n_2)^2}{4n_1n_2} = 10 \log \frac{2.5^2}{6} = 0.18 \text{ dB}$$

Because two such boundaries exist, the total Fresnel loss is approximately 0.35 dB per connection. Index-matching fluids or epoxy can reduce this loss term to near zero.

A second mismatch loss arises when light is coupled from a large-diameter core to a smaller-diameter receiving core. The resulting area mismatch loss  $\alpha_A$  is

$$\alpha_A = 20 \log \frac{\text{larger core diameter}}{\text{smaller core diameter}} \quad \text{dB} \quad (21.17)$$

If the receiving fiber core is larger than the transmitting core, the area mismatch loss is zero.

**example 21.9** A 200- $\mu\text{m}$  fiber core couples optical power from an LED into a 50- $\mu\text{m}$  graded-index fiber core. Calculate the area mismatch coupling loss.



**solution** Since the sending fiber is larger than the receiving fiber, the area mismatch coupling loss is

$$\begin{aligned}\alpha_A &= 20 \log \frac{200 \mu\text{m}}{50 \mu\text{m}} \text{ dB} \\ &= 12 \text{ dB}\end{aligned}$$

A third mismatch loss term occurs if the sending core numerical aperture is larger than that for the receiving fiber core. This loss is given by

$$\alpha_{NA} = 20 \log \frac{\text{sending core N.A.}}{\text{receiving core N.A.}} \quad \text{dB} \quad (21.18)$$

If the sending core N.A. is less than the receiving core N.A., then  $\alpha_{NA} = 0$ .

**example 21.10** For the two fibers in Example 21.9, the 200- $\mu\text{m}$  sending fiber has N.A. = 0.50, and the 50- $\mu\text{m}$  receiving fiber has N.A. = 0.20. What is the coupling loss owing to mismatch of numerical apertures?

**solution** From Eq. (21.18) the coupling loss owing to mismatch of the numerical apertures is

$$\begin{aligned}\alpha_{N.A.} &= 20 \log \frac{0.50}{0.20} \text{ dB} \\ &= 8 \text{ dB}\end{aligned}$$

The total loss in decibels experienced in connecting is the sum of the effects resulting from the misalignment, fiber end separation, and mismatch terms. By combining the results of Examples 21.9 and 21.10 with Fresnel and alignment losses, the total connecting attenuation is

Connecting alignment	2 dB
$\alpha_{N.A.}$	8 dB
$\alpha_A$	12 dB
Fresnel loss	0.3 dB
Total loss	22.3 dB

**example 21.11** An LED couples 700  $\mu\text{W}$  of optical power into a 200- $\mu\text{m}$ -diameter, 0.50-NA fiber pigtail. This fiber is connected to a 50- $\mu\text{m}$ -diameter, 0.20-N.A. telecommunications fiber. (See Example 21.10.) What is the power coupled into the optical transmission line?

**solution**

$$\begin{aligned}P_c &= P_o \times \text{fraction lost} \\ \text{or} \quad P_c (\text{dB}_m) &= P_o (\text{dB}_m) - \text{losses (dB)}\end{aligned}$$

From Example 21.10 the coupling loss is 22.3 dB, so

$$\begin{aligned}P_c &= -1.5 \text{ dB}_m - 22.3 \text{ dB} \\ &= -23.8 \text{ dB}_m\end{aligned}$$

By converting from  $\text{dB}_m$  to milliwatts, the coupled power is only

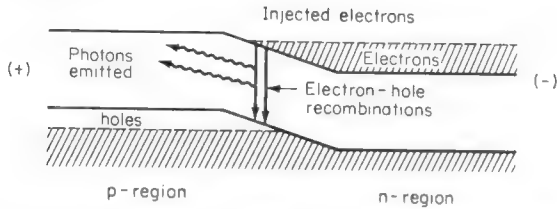
$$\begin{aligned}P_c &= 10^{-23.8/10} \text{ mW} \\ &= 0.042 \text{ mW} = 4.2 \mu\text{W}\end{aligned}$$

It should not be inferred that the manufacturer has an inferior product; the laws of physics apply to cause the coupling loss.

### 21.3. OPTICAL TRANSMITTERS AND SOURCES

The two types of optical sources which have been used widely are the light-emitting diode (LED) and the injection laser diode (ILD). These are solid-state pn junction devices that produce an increased light output with an increased forward-bias current. Electrons injected across the forward-biased pn junction combine with holes to produce photons of light which escape from the structure, as illustrated in Fig. 21.24. Various semiconductor compounds of gallium (Ga), arsenic (As), indium (In), and phosphorus (P) are used to produce devices which have a peak light emission at wavelengths from 600 nm in the visible part of the optical spectrum to 1600 nm in the infrared part. The range of light emission in the optical spectrum for each material type is given in Fig. 21.25. Included also are other gas and solid-

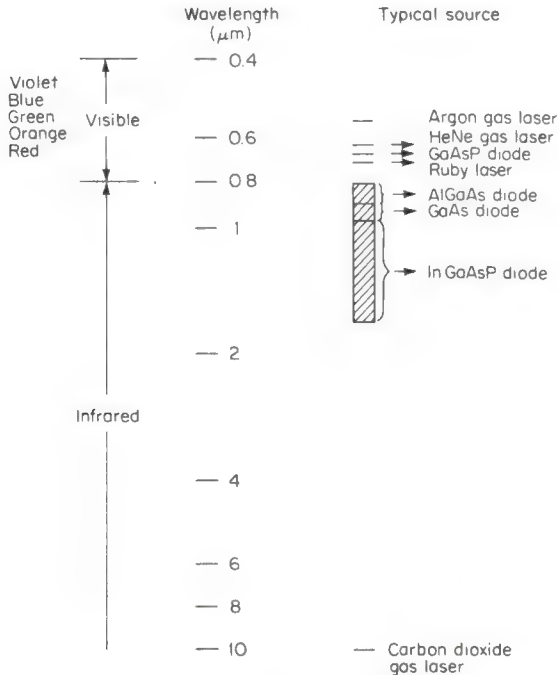




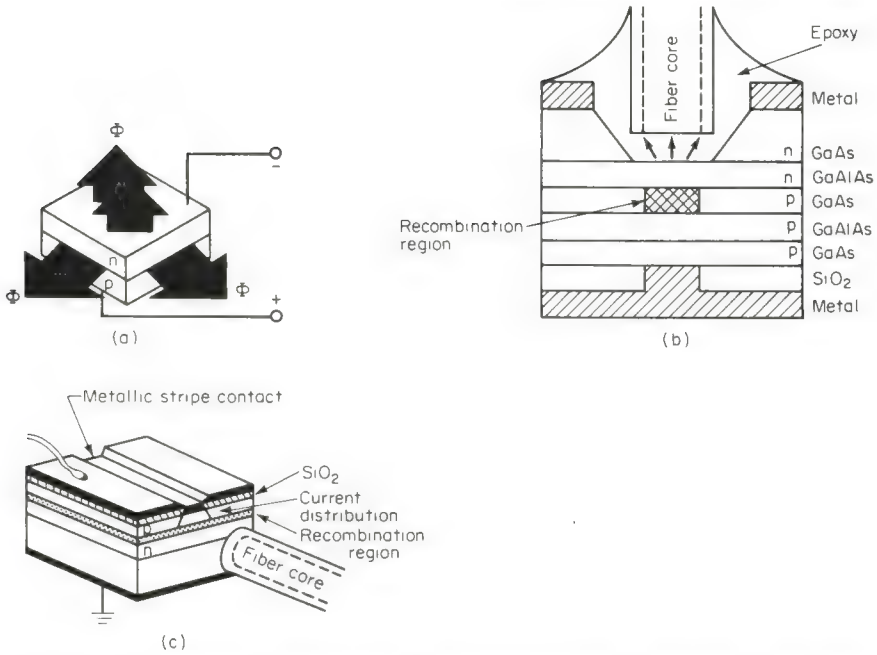
**Fig. 21.24** Energy band diagram for a semiconductor light-emitting diode. Electrons injected into the conduction band of the n region recombine with holes in the valence band to produce photons of light. The larger the electron injection, the greater the production of photons.

state lasers used in optical communications. Many installations presently use aluminum gallium arsenide (AlGaAs) LEDs and ILDs which emit in the range of 780 to 900 nm. Future wide-bandwidth, long-distance links are expected to incorporate the InGaAsP devices, which transmit between 920 and 1650 nm.

The choice of an LED or ILD optical source rests on understanding the differences between the devices. For example, the ILD has a shorter risetime than the LED, which makes the ILD more suitable for wide-bandwidth and high-data-rate applications. In addition, more optical power can be coupled into a fiber with an ILD, which is important for long-distance transmission. However, the laser diode is strongly temperature-dependent and has a significantly higher cost. Which device type, LED or ILD, and what carrier wavelength to select depend on the transmission distance, data rate or bandwidth, fiber type, detector, and total system cost.



**Fig. 21.25** The optical portion of the RF spectrum and typical sources at the various wavelengths. Solid-state, gas, and semiconductor optical emitters are given.



**Fig. 21.26** (a) Directions of light emission for a light-emitting diode. (b) Structure and fiber coupling scheme for a heterojunction gallium aluminum arsenide surface emitting LED. (c) Structure and fiber alignment for an edge emitting LED.

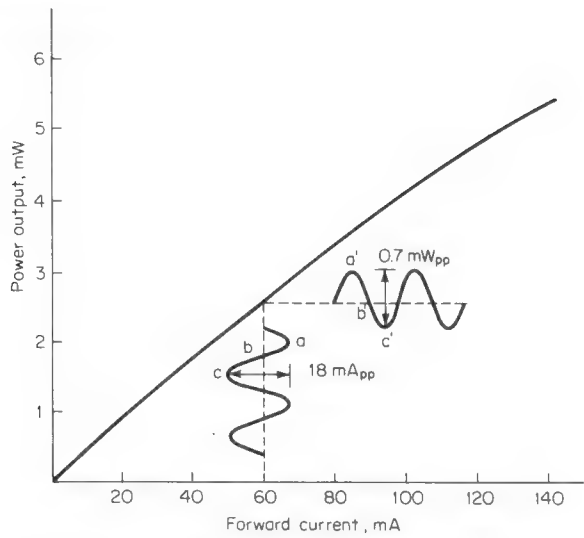
**LIGHT-EMITTING DIODES** The two most common LED structures are the surface emitter and the edge emitter, shown in Fig. 21.26. Surface emitters have a higher light output and efficiency than edge emitters because the light produced at the pn junction travels through only a thin absorbing layer of semiconductor material before it reaches the surface. For edge emitters, however, the photons travel the length of the structure parallel to the surface. Hence, more light is lost through absorption by the semiconductor material. Epitaxial layers of compound semiconductor materials form the p- and n-type regions. A heterojunction occurs if the pn junction is between two different types of semiconducting compounds, such as GaAs and AlGaAs.

The optical output characteristic versus current for a typical LED is illustrated in Fig. 21.27. For a drive current of 100 mA, the total available optical power for commercially available devices ranges between 1 and 10 mW. Amplitude modulation of the diode forward current produces a change in optical output intensity, as diagrammed in Fig. 21.28. The maximum modulation index for analog transmission is typically on the order of 80 percent; for digital transmission, it is 100 percent. The output power for a given bias current may degrade up to 3 dB over the lifetime of the device, which for AlGaAs LEDs may exceed 100,000. Driver circuits for digital and analog modulation are found in Figs. 21.28 and 21.29.

Light-emitting diodes present the optical engineer or technician with several problems. One problem is that emitted light is spatially distributed as shown in Fig. 21.30. The optical intensity (power per unit area) at an angle  $\theta$  from the normal to the emitting surface is given by

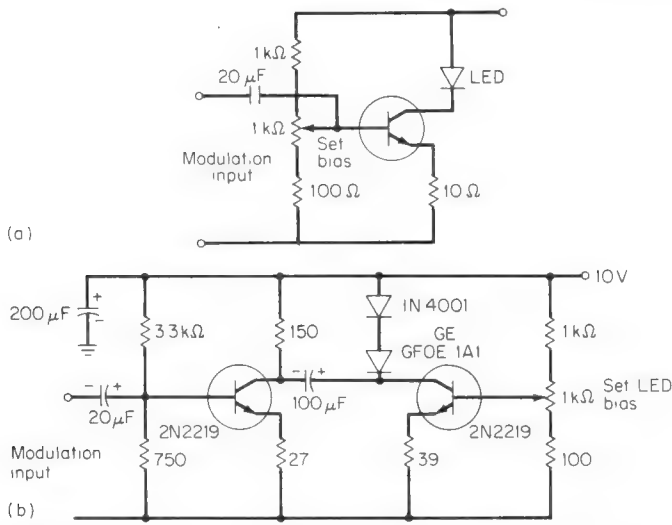
$$I(\theta) = I_0 \cos^m \theta \quad (21.19)$$

where  $I_0$  is the intensity measured normal to the surface and  $m$  is an integer which describes how well the pattern is collimated. For  $m = 1$ , the output pattern is called a *Lambertian distribution*. For edge emitters,  $m = 2$  or 3; for ILDs,  $m = 3$  or 4. If the fiber is located too far from the emitting area, much of the light misses the fiber core.

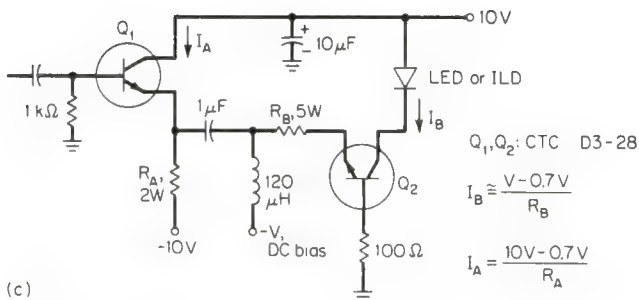


**Fig. 21.27** Optical output power as a function of dc forward-bias current for a typical LED. A 9-mA peak modulation current applied at a 60-mA dc bias current produces a 14 percent modulation of the optical output power.

To minimize the loss of light power, a number of coupling schemes are used, as illustrated in Figs. 21.31 and 21.32. Devices with an attached optical fiber are identified as “pigtailed.” Even with the most efficient coupling schemes, however, the power coupled into the optical fiber is 10 to 20 dB below that emitted by the LED. The amount of coupled power depends on the type of optical fiber as well as in the coupling scheme. How to determine the coupled power is discussed in the section on optical fibers.

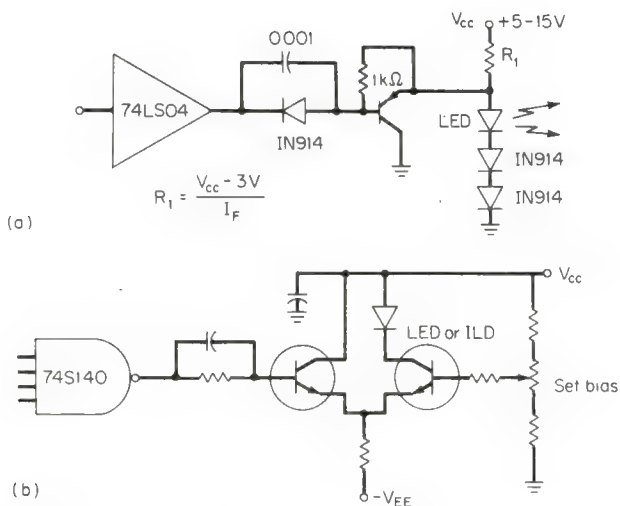


**Fig. 21.28** (a) A basic bias and modulation driver for an LED. (b) A better LED bias and driver circuit which has a 100-kHz bandwidth. (c) An LED bias and driver circuit with a 50-MHz bandwidth.



(c)

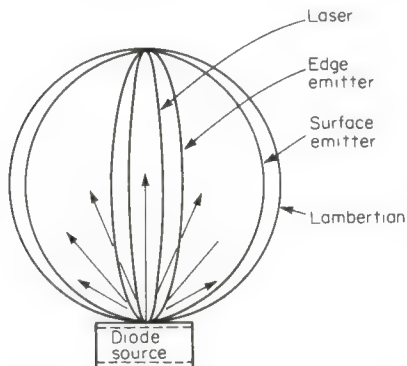
**FIG. 21.28** (Cont.)



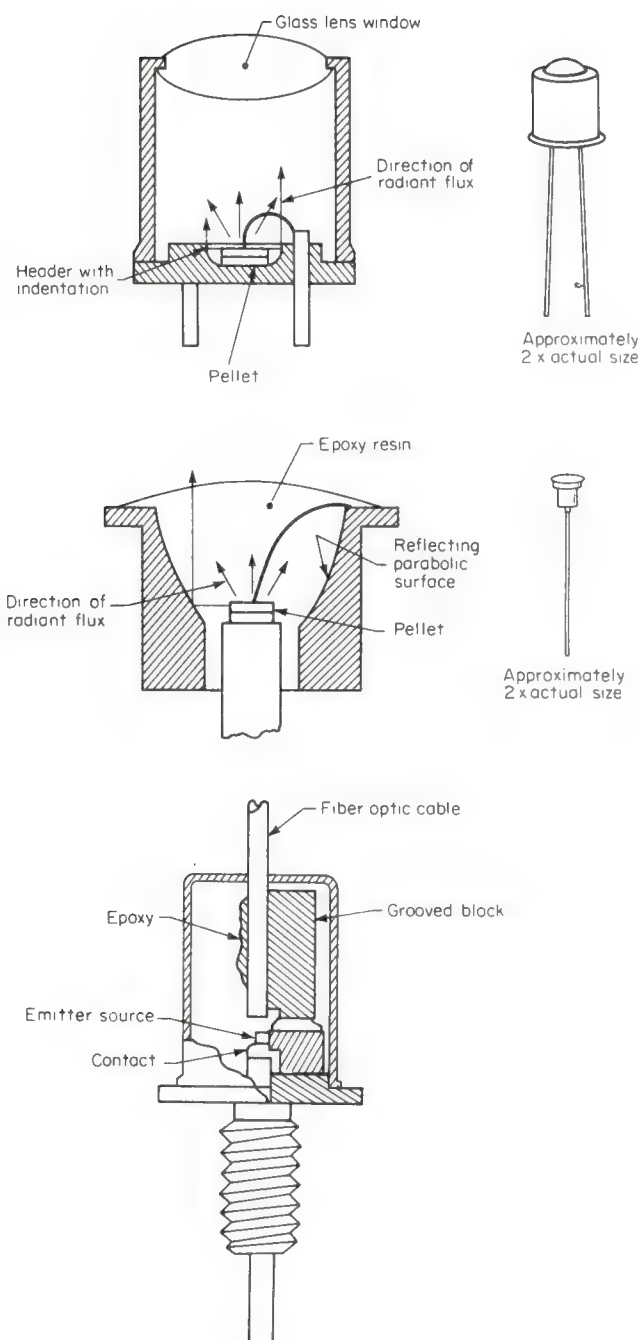
(a)

(b)

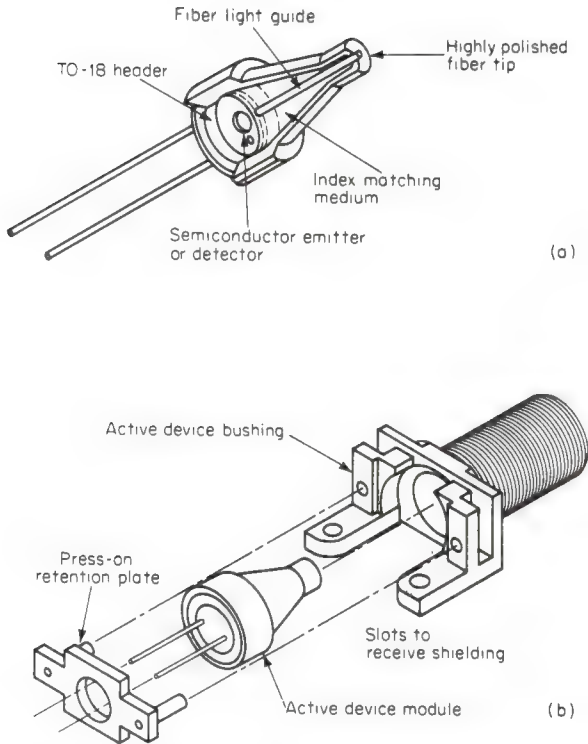
**Fig. 21.29** Examples of digital driver circuits for digital applications (a) below 100 kilobits per second and (b) below 50 megabits per second.



**Fig. 21.30** Spatial emission pattern for light-emitting and laser diodes which shows the intensity as a function of angle from the normal to the emitting surface. A laser emits light within a narrow cone which results in efficient coupling of light into a fiber core.



**Fig. 21.31** Mounting packages for light-emitting and laser diode pellets. (Courtesy RCA, Solid State Division, Electro Optics and Devices, Lancaster, Pa.)

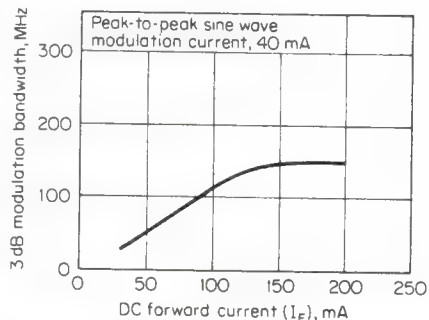


**Fig. 21.32** (a) Ferruled LED package and (b) the mounting bushing for use with the AMP connector illustrated in Fig. 21.20. (Courtesy Motorola Semiconductor Products, Inc., Phoenix, Ariz., and AMP, Inc., Harrisburg, Pa.)

A limitation of the LED is a relatively low-modulation bandwidth. GaAs and GaAsP devices with a risetime in the range of 600 to 800 ns have a bandwidth of 500 kHz or less. In contrast, edge-emitting AlGaAs diodes have risetimes between 2 and 20 ns and bandwidths greater than 200 MHz. For surface emitters, the inherent junction capacitance limits the maximum modulation bandwidth to less than 30 MHz. The modulation bandwidth varies with dc forward-bias current, as shown in Fig. 21.33.

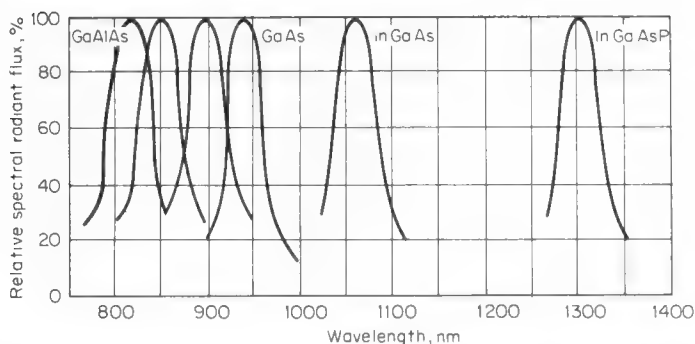
The third problem is that the optical output power is spread over a band of wavelengths, as shown in Fig. 21.34. It is defined as the full width at half maximum (or FWHM), which is approximately 40 nm for most devices. Light rays with different wavelengths may travel down an optical guide with different velocities. Hence, a light pulse coupled into an optical fiber from an LED is broadened. The amount of pulse broadening plays a role in limiting the bandwidth of an optical link.

The characteristics of several LED com-



**Fig. 21.33** Modulation bandwidth as a function of dc forward-bias current for an LED. (Courtesy RCA, Solid State Division, Electro Optics and Devices, Lancaster, Pa.)





**Fig. 21.34** Typical spectral emission characteristics of light-emitting diodes fabricated from different semiconductor materials. (Courtesy RCA, Solid State Division, Electro Optics and Devices, Lancaster, Pa.)

ponents and transmitters are listed in Table 21.3. It is important that the selection of the optical emitter be made to match optimum performance properties of the fiber to be used as well as the maximum response of the photodetector.

**INJECTION LASER DIODES** A laser diode is similar in structure to that for a heterojunction edge-emitting LED except that the end surfaces are cleaved or polished to produce a mirror finish. (Refer to Fig. 21.26c.) In addition, a channel or stripe geometry is used to

**TABLE 21.3 LED Components and Transmitters**

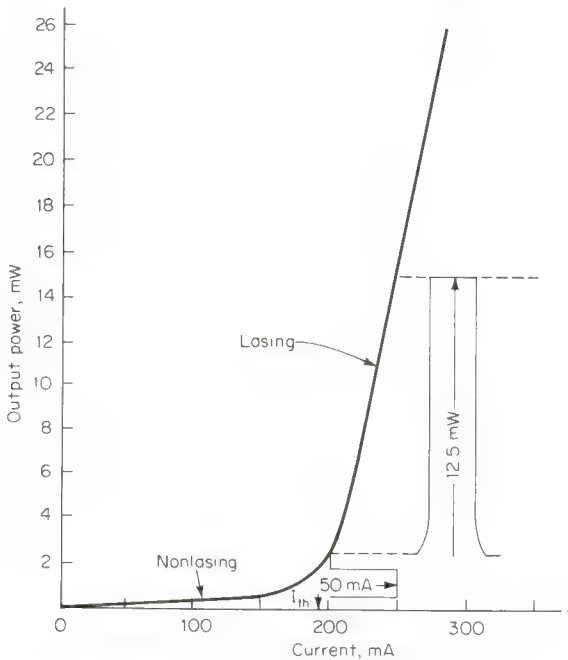
Wavelength (nm)	Bandwidth (MHz)	Power ( $\mu$ W)	Port/pigtail [diameter ( $\mu$ m)/N.A.]	Manufacturer and model
Light-emitting diodes				
820	32	240	125/0.48	Laser Diode IRE 160FA
900	18	140	200/0.48	Motorola MFOE102F
820	38	700	200/0.50	Motorola MFOE106F
850	45	200	125/0.25	NEC NDL 4103 CP
820	150	60	100/0.21	RCA C86008E
1200	90	15	50/0.20	ITT T7632
1270	90	20	50/0.21	Laser Diode DE-1000F
Analog LED transmitters				
820	30	140	300/0.22	LeCroy FAT-4H
820	20	20	62/0.22	MERET MDL 276-TV
840	20	100	—	ITT T 6211
Digital LED transmitters				
	Bit rate (Mb/s)			
670	2	15		Burr Brown 3713T
660	5	15	370/0.53	LeCroy Diplink 1
820	10	100	100/0.30	Hewlett-Packard HFBR-1002
820	40	200	55/0.22	Laser Diode LDT-256
820	50	160	200/0.22	LeCroy FDT-5L
820	20	40	63/0.21	RCA C 86012E
820	10	275	200/0.48	Spectronics SPX 4140

reduce the region where electron injection occurs at the junction. This area reduction increases the injection current at the junction and increases the photon output. The mirrored end faces cause a large fraction of the light produced at the junction to be reflected into the channel region. Produced is an optical waveguide cavity which guides the photons to stimulate the generation of additional photons in phase with the stimulus. The result is light amplification by stimulated emission of radiation, from which the acronym *LASER* is obtained.

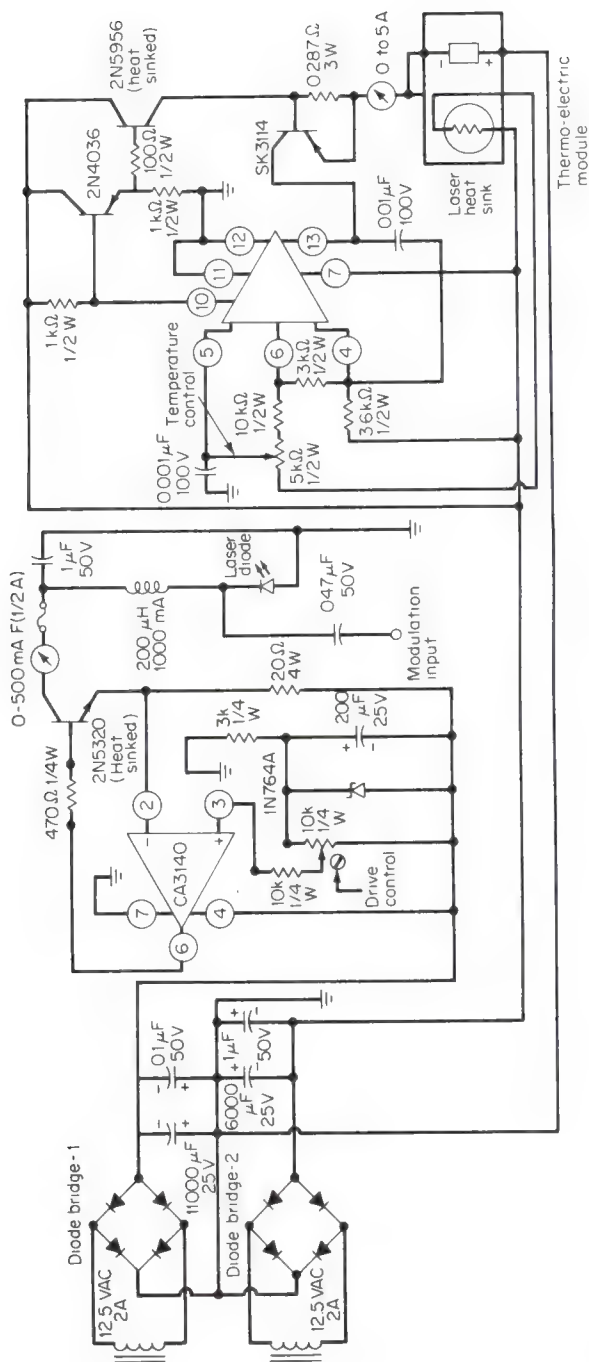
The output characteristic versus drive current for an ILD differs from that for the LED, as illustrated in Fig. 21.35. Significant light output does not occur unless the drive current exceeds a threshold; spontaneous emission similar to that for an LED occurs below threshold. Because small changes in drive current above the device threshold produce substantial changes in output optical power, the ILD is well suited for applications in digital communications systems. An example of a dc bias and modulation circuit for ILDs is given in Fig. 21.36.

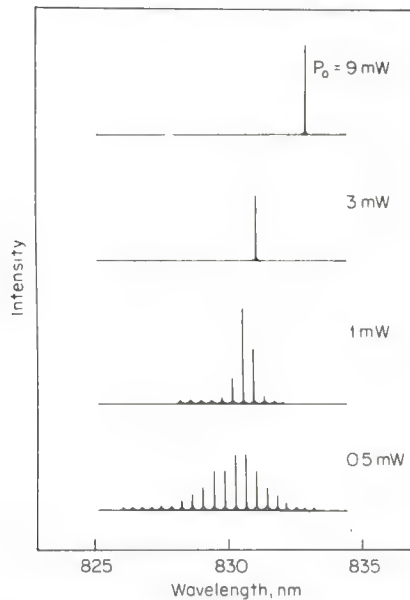
A typical emission spectrum is shown in Figs. 21.37 and 21.38. At low power levels, the output power is distributed among several narrow emission lines or modes. As the drive current and optical power are increased, the power shifts to a dominant mode which is only a few nanometers wide. Note that the mode pattern and dominant emission wavelength are a function of the drive current. Such multimode operation contributes to high-frequency noise. Single-mode lasers have been developed that do not suffer from these characteristics.

The ILD has three main advantages over the LED. The first is its directionality. Because of the stimulated emission, the light emerges in a narrow beam, as indicated in Fig. 21.30. Hence, a large fraction of the optical power may be coupled into a small fiber core. Typically, the coupling loss is less than 3 dB. The second advantage is the very low junction



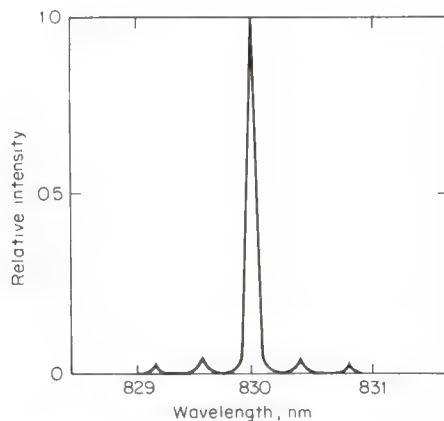
**Fig. 21.35** Optical output power as a function of dc forward-bias current for an injection laser diode. Stimulated emission of light does not occur unless the threshold current is exceeded. Below threshold, spontaneous light emission occurs similar to that for an LED. The graph illustrates how a 50-mA current pulse yields a 12.5-mW change in the optical output power.





**Fig. 21.37** Typical spectral emission characteristic for an aluminum gallium arsenide (AlGaAs) laser diode as a function of optical output power. At low power levels, multimode emission occurs.

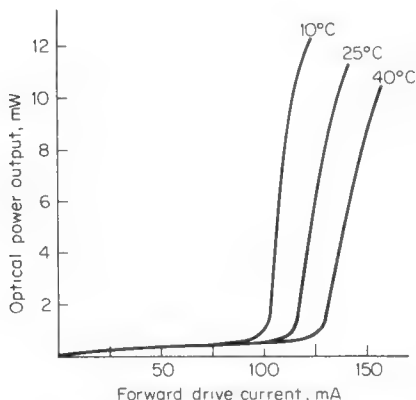
capacitance, which permits a high-modulation bandwidth. Present AlGaAs devices are available with bandwidths exceeding 1 GHz. The third advantage is the narrow wavelength spread in the output spectrum, typically on the order of 1 to 2 nm. The result is a much smaller material dispersion or pulse broadening. For these reasons laser sources can be used to transmit data at higher rates over a given distance than is possible with LEDs.



**Fig. 21.38** Typical spectral characteristic of a single-mode laser emission line. The line width typically is less than 0.5 nm, which reduces the effects of material dispersion.

A disadvantage of the ILD is the strong temperature dependence of the output characteristic curve, illustrated in Fig. 21.39. To control this dependence, special optical and electronic feedback drive circuits often are required. However, the result is a more complicated and costly transmitter. ILD sources may be either obtained as individual components or packaged as a complete transmitter which includes all driver feedback circuits. Several ILD components and transmitters are listed in Table 21.4.

Lasers also exhibit temporal instabilities which lead to "noise" in the optical output. One type of fluctuation occurs when changes in fiber-interface reflections perturb the laser optical



**Fig. 21.39** Optical output power as a function of dc forward-bias current for a laser diode at different operating temperatures. A laser biased for proper operation at 25°C may cease to operate at 40°C. Electronic regulation of the device temperature usually is required, as shown in Fig. 21.36.

cavity. A second noise category occurs as a train of discrete, narrow optical pulses even when the drive current is not modulated. These self-oscillations occur at frequencies above 200 MHz. A third category is modal noise, which arises from changes in the distribution of light among the fiber propagation modes. Such changes can occur as a result of wavelength changes when the laser is modulated. These noise characteristics have limited the application of the laser in broadband services such as CATV, which require high signal-to-noise ratios.

## 21.4 OPTICAL DETECTORS AND RECEIVERS

The optical receiver must detect the received optical power and amplify the resulting electric signal with minimal distortion. It consists of a photodetector and a preamplifier, the choice of which greatly affects how well this requirement is satisfied. To appraise the performance of a receiver, it is necessary to evaluate the bandwidth,

gain, noise, spectral response, sensitivity, and dynamic range, all of which affect the cost of a receiver system.

The two general classes of solid-state photodetectors are the reverse-biased pn diode and the phototransistor. One common type of diode has a PIN structure, illustrated in Fig. 21.40. A wide intrinsic region is placed between the p- and n-type materials to increase the junction width, which lowers capacitance and lowers bulk leakage currents that contribute to noise. Photons of light incident on the diode junction generate electron-hole pairs in the depletion region to produce an electric current proportional to the received optical power. The diode current ranges from picoamperes ( $10^{-12}$  A) to a few tenths of milliamperes. The efficiency of photon-to-carrier conversion is typically between 85 and 95 percent. Good efficiency and response times are obtained with reverse-bias voltages between 10 and 50 V.

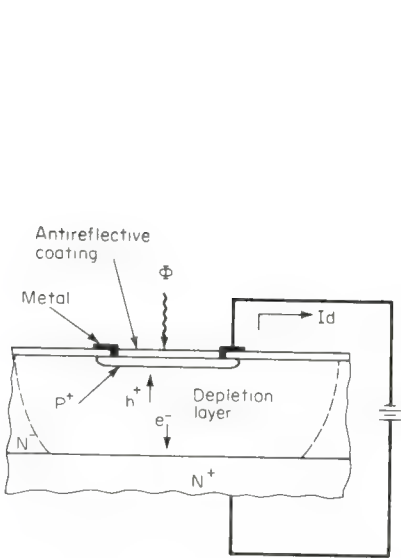
Another diode detector type is the avalanche photodiode (APD) in which the photoexcited electrons trigger an avalanche of electrons through a high reverse-bias electric field. (Refer to Fig. 21.41.) This effect yields an apparent conversion gain and gives the detector its name. For this reason the APD is 10 to 15 dB more sensitive than the PIN diode. However, it is costly to produce, requires a reverse bias of 150 to 450 V, and exhibits a significant temperature dependence in the photoconversion gain. For many applications, these characteristics make APD devices less suitable than the PIN photodiodes.

In the phototransistor, the light detection diode becomes the base-collector junction of the transistor. (See Fig. 21.42.) The junction area is increased to improve the photon capture, but this causes higher base-collector capacitance and increases the Miller effect capacitance. Consequently, phototransistors are limited to applications requiring bandwidths less than 500 kHz. However, they can be manufactured at a cost lower than that for PIN diodes. A typical response efficiency is 80 percent for silicon devices for 850-nm light.

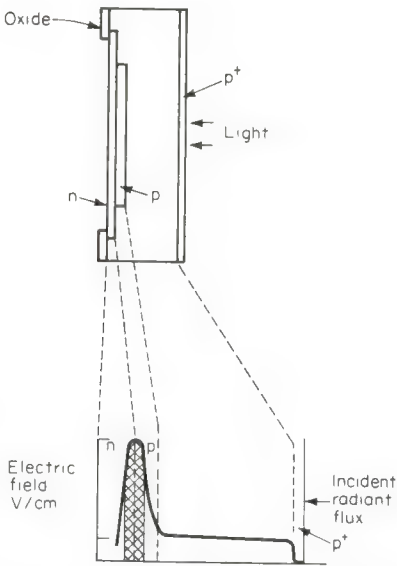
To amplify the small photocurrent, a low-noise preamplifier is required. Two commonly used receiver configurations are the transimpedance amplifier and the voltage amplifier,

**TABLE 21.4 Laser Diodes and Transmitters**

Wavelength (nm)	Bandwidth (MHz)	Power ( $\mu$ W)	Port/pigtail [diameter ( $\mu$ m)/N.A.]	Manufacturer and model
Lasers				
780-850	>1200	7000	—	General Optronics GOL5
800-890	2000	1500	55/0.22	Laser Diode LCW-10F
840	500	500	55/0.25	ITT T 912
830	1000	2500	63/0.21	NEC NDL 3108 P
820	1000	2000	100/0.21	RCA C 86006 E
1300	1000	1000	50/0.20	RCA 86022 E
Analog ILD transmitters				
780-850	1250	3000	55/0.25	General Optronics GO-ANA
830	500	1000	—	Times Fiber OTL-1101
840	150	2000	—	ITT T-6247
1300	500	3000	50/0.20	General Optronics GO-ANA
Digital ILD transmitters				
	Bit rate (Mb/s)			
780-850	275	>1000	—	OIS OTX-5100
840	300	1000	—	ITT T6146
820	150	500	—	Meret MDL-4595



**Fig. 21.40** Structure of a PIN photodiode which shows the generation of a hole-electron pair by the absorption of a photon. The applied reverse bias separates the carriers to yield the diode current. (Courtesy RCA, Solid State Division, Electro Optics and Devices, Lancaster, Pa.)



**Fig. 21.41** Structure and electric field distribution for an avalanche photodiode. The large electric field causes impact ionization of additional carriers which multiplies the photogenerated carriers. (Courtesy RCA, Solid State Division, Electro Optics and Devices, Lancaster, Pa.)



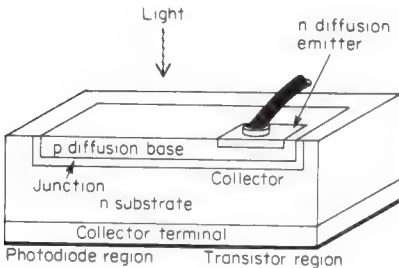
shown in Fig. 21.43. The transimpedance amplifier is normally used for bandwidths below 100 MHz for which the output voltage is given by

$$V_{out} = -I_d R_f \quad (21.20)$$

where  $R_f$  is the feedback resistance. It is attractive from a design standpoint because it provides better input and output impedance stability and yields reduced distortion. The voltage amplifier, however, is simpler and less costly to implement for large-bandwidth applications. It permits a larger bandwidth and better signal-to-noise ratio (SNR). However, because the RC network of the diode and load resistance acts as a differentiator, an output equalizer network is required. The RC product of the input circuit determines the amplifier bandwidth (BW), given by

$$BW = \frac{1}{2\pi RC} \quad (21.21)$$

where  $R$  is the total input capacitance and  $C$  is the total device and amplifier input capacitance. For both PIN and APD photodetectors, the risetime typically is less than 3 ns. Hence, the bandwidth is limited usually by the preamplifier circuit.



**Fig. 21.42** Structure of a phototransistor. (Courtesy General Electric Company, Semiconductor Products Department, Auburn, N.Y.)

**example 21.12** A transimpedance amplifier has a feedback resistance of 20 kΩ and a feedback capacitance of 10 pF. If the PIN photodiode capacitance is 5 pF, calculate the bandwidth.

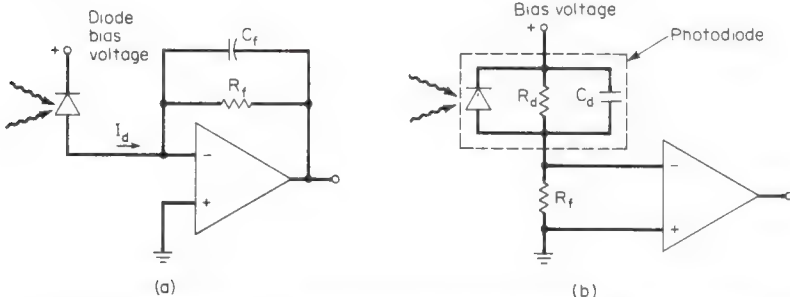
**solution** From Eq. (21.21) the bandwidth is

$$BW = \frac{1}{2\pi(20 \text{ k}\Omega)(10 \text{ pF} + 5 \text{ pF})} = 530 \text{ kHz}$$

Just as in any communications system, the receiver configuration must detect and amplify the received signal with minimal distortion. The SNR must be great enough to maintain signal integrity and to keep

errors below an acceptable minimum. The choice of detector and preamplifier circuit affects how well this is accomplished. For either amplifier configuration, FET stages yield lower noise figures below 10 MHz; above 10 MHz, bipolar devices are better. Which configuration to select depends on the circuit cost, required bandwidth, and available SNR. The detector and receiver specifications which determine the performance characteristics are described below.

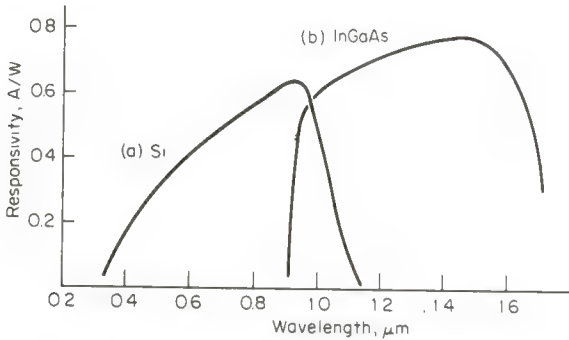
**RESPONSIVITY** The optical-to-electrical conversion of a photodetector is given by the responsivity  $R_d$ :



**Fig. 21.43** Diagrams for (a) a transimpedance amplifier and for (b) a voltage amplifier used in optical receivers.

$$R_d = \frac{\text{photocurrent}}{\text{received optical power}} \quad (21.22)$$

The responsivity depends on the diode type, semiconductor material, bias voltage, and wavelength of the optical carrier. For the range of 650 to 950 nm, silicon photodiodes are used; for longer wavelengths the newly developed indium gallium arsenide (InGaAs) detectors are used. Typical responsivities for PIN detectors are shown in Fig. 21.44 and listed in Table 21.5. Peak values are 0.4 to 0.7 A/W for PIN diodes and 75 A/W for APD detectors.



**Fig. 21.44** Responsivity as a function of a wavelength for (a) silicon and (b) indium gallium arsenide PIN photodiodes.

The 100-fold increase in responsivity for the APD arises from the avalanche process. Germanium (Ge) detectors also may be used in the spectral range of 800 to 1800 nm; however, a large reverse current produces a high noise level.

An optical receiver also may have a responsivity defined in terms of the output voltage:

$$R_{\text{amp}} = \frac{\text{output voltage}}{\text{received optical power}} \quad (21.23)$$

For a transimpedance amplifier this equation may be rewritten by means of Eqs. (21.22) and (21.23) in the form

$$R_{\text{amp}} = R_d R_f \quad (21.24)$$

**example 21.13** Suppose the amplifier circuit in Fig. 21.43a has  $R_f = 20 \text{ k}\Omega$ . For a PIN detector responsivity of 0.60 A/W, determine the amplifier responsivity.

**TABLE 21.5** PIN and APD Detectors

$\lambda_0$ (nm)	$R$ (A/W)	Risetime (ns)	NEP ( $10^{-13} \text{ W}/\sqrt{\text{Hz}}$ )	Typical bias (V)	Manufacturer and model
PIN diodes					
830	0.6	3	1.0	-45	RCA C30920E
820	0.3	3		-5	Spectronics SD-3322
900	0.4	25	0.05	-20	Motorola MFOD102F
940-1250	0.6	2	5	-5	ITT T7731
900-1650	0.6	1	5	-5	RCA C30979E
APD diodes					
830	77	0.5	0.3	-225	RCA C30921E
800-900	85	1.2	0.08	-300	Centronic APD 05-4R

**solution**

$$\begin{aligned}
 R_{\text{amp}} &= R_f R_d \\
 &= (20 \text{ k}\Omega) (0.6 \text{ A/W}) \\
 &= 12\,000 \text{ V/W}
 \end{aligned}$$

A number of manufacturers supply integrated photodetector and preamplifier packages. (Refer to Tables 21.6 and 21.7.)

Because the photocurrent is proportional to the received optical power, care must be exercised to interpret level changes and signal-to-signal noise values. Recall that the decibel is given by

$$\text{dB} = 10 \log \frac{P}{P_{\text{ref}}} \quad (21.25)$$

or, in terms of a current ratio, as

$$\text{dB} = 20 \log \frac{I}{I_o} \quad (21.26)$$

Therefore, a 50 percent decrease in *optical* power ( $-3\text{-dB}$  change) produces a  $-6\text{-dB}$  change in the output electric current. Hence, the optical and electric signal properties are related by

$$\text{Optical SNR (dB)} = \frac{1}{2}(\text{electric SNR}) \text{ (dB)} \quad (21.27)$$

This optical power-to-current conversion affects the interpretation of the bandwidth of an optical system. It is defined as the modulation frequency where the received optical power decreases by 3 dB. At this frequency, the electric output is down by 6 dB, however. Consequently, the electrical and optical bandwidths are related by

$$\text{Electrical bandwidth} = 0.707(\text{optical bandwidth}) \quad (21.28)$$

**example 21.14** If an optical fiber has an optical bandwidth-length product of  $600 \text{ MHz} \cdot \text{km}$ , what is the equivalent electrical bandwidth-length product?

**solution** By means of Eq. (21.28),

$$\begin{aligned}
 \text{Electrical BW} &= 0.707(600 \text{ MHz} \cdot \text{km}) \\
 &= 424 \text{ MHz} \cdot \text{km}
 \end{aligned}$$

Failure to recognize the difference in the optical and electrical bandwidths easily may lead to a disappearance of system bandwidth.

**NOISE EQUIVALENT POWER** In contrast to the normal communications receiver, the noise in the output of an optical receiver originates from two sources: the detector/amplifier circuit, and shot noise resulting from the random arrival of photons at the detector surface. (Photon-generated shot noise is sometimes called quantum noise.) This causes the noise level at the preamplifier output to depend on the level of the received optical power and the

**TABLE 21.6 Fiber-Optic Receivers**

Wavelength peak (nm)	Bandwidth (MHz)	TNEP ( $\mu\text{W}$ )	Manufacturer and model
PIN receivers			
820	25	0.065	LeCroy FAR-4HS
820	5	0.5	Fibronics ALS-R1
840	6	1	ITT T 6358A
900	35	2.0	Motorola MFOD403F
900	10	0.3	Motorola MFOD404F
APD receivers			
830	300	—	Times Fiber OR-2111A
840	6	0.06	ITT T 6358B
890	130	—	EG & G ODL-4A
905	1–100	0.02	Meret MDL-259

TABLE 21.7 Digital Fiber-Optic Receivers

Wavelength peak (nm)	Bit rate (Mb/s)	Sensitivity ( $\mu\text{W}$ )	Bit error rate ( $\text{s}^{-1}$ )	Manufacturer and model
PIN receivers				
650–1000	0.25	0.015	$10^{-9}$	Burr-Brown 3713R
670	1.0	0.25	$10^{-8}$	LeCroy DipLink 1
700–840	10	0.8	$10^{-9}$	Hewlett-Packard HFBR-2001
780–850	0.1–50 (150)	1 (5)	$10^{-9}$	OIS ORX-5000 (ORX-5100)
800–900	20	1.5	$10^{-8}$	LeCroy FDR-5AL
820	20 NRZ	0.5	$10^{-9}$	RCA C86012E
840	300	0.3	$10^{-8}$	ITT T6092
820	0.01–10	0.7	$10^{-9}$	Spectronics SPX 4141
APD receivers				
780–850	0.1–50 (150)	0.05 (0.2)	$10^{-9}$	OIS ORX-5010 (5110)
840	107	0.01	$10^{-8}$	ITT T6196
905	100	0.5	$10^{-12}$	Meret MDL 4988

required SNR. The total output noise may be specified in terms of the total noise equivalent power (TNEP):

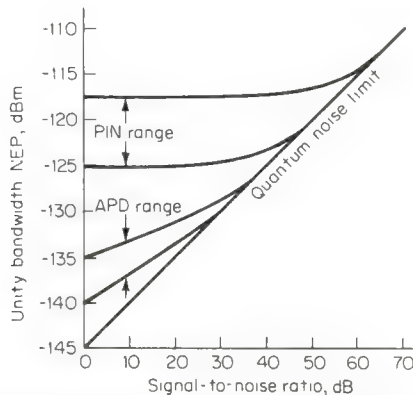
$$\text{TNEP} = \frac{\text{output noise current or voltage}}{\text{responsivity}} \quad (21.29)$$

It represents the minimum detectable optical signal ( $\text{SNR} = 1$ ). For example, the RCA C30807 PIN diode with a dark current of 10 nA and  $R_d$  of 0.6 A/W has a TNEP of 16.7 nW.

Recall that the rms noise voltage is proportional to the square root of the bandwidth  $BW$ . By dividing the TNEP by  $\sqrt{BW}$ , the noise equivalent power (NEP) is obtained:

$$\text{NEP} = \frac{\text{TNEP}}{\sqrt{BW}} \quad (21.30)$$

It represents the noise in a unity bandwidth and provides a basis for comparing noise levels of different receivers, as in Fig. 21.45. At low optical intensities (low SNR), circuit noise



**Fig. 21.45** Range of noise equivalent power (NEP) at unity bandwidth as a function of electric signal-to-noise ratio for optimum PIN and avalanche photodiode receivers. Actual receiver circuits may have higher NEP values than that indicated.

dominates the output of a PIN receiver but not an APD receiver; that is, the APD is limited by quantum noise. For this reason the APD detector is preferred for long-distance transmission of digital information since the required SNR rarely exceeds 20 dB. However, for many analog transmissions with high SNR values, shot noise dominates, so the PIN diode is the preferred detector in these applications.

**SENSITIVITY** The sensitivity represents the optical power at the receiver input which yields a given electric SNR at the output. In terms of the TNEP or NEP, it is

$$\begin{aligned} P_r (\text{dB}_m) &= \text{TNEP} (\text{dB}_m) + \frac{1}{2} \text{SNR} (\text{dB}) - 10 \log m \\ &= \text{NEP} (\text{dB}_m) + 10 \log B + \frac{1}{2} \text{SNR} (\text{dB}) - 10 \log m \end{aligned} \quad (21.31)$$

where  $m$  is the modulation index of the optical carrier. Since the optical power level is the desired quantity, one-half of the electric SNR is used. The value of the NEP may be read from one of the curves in Fig. 21.45. Two examples given below illustrate the calculation of the sensitivity.

**example 21.15** A PIN diode receiver is used as part of a TV remoting link. The TV signal modulates the optical carrier with a modulation index of 0.7, a bandwidth of 4.5 MHz, and an SNR of 50 dB. What is the required receiver sensitivity?

**solution** From Fig. 21.45 the NEP for a typical low-noise receiver for this application is approximately  $-118 \text{ dB}_m$ . The sensitivity is computed from Eq. (21.31):

$$\begin{aligned} P_r &= -118 \text{ dB}_m + 10 \log(4.5 \text{ MHz}) + 25 \text{ dB} - 10 \log 0.7 \\ &= -28 \text{ dB}_m = 1.6 \mu\text{W} \end{aligned}$$

**example 21.16** The same TV signal in Example 21.15 is digitized at the Nyquist rate (9 MHz) with an NRZ, 8-bit code. The bit rate is 72 Mb/s, which requires a minimum bandwidth of 36 MHz. What is the required sensitivity of a PIN receiver and of an APD receiver?

**solution** For a bit error rate of  $10^{-9}$ , the electrical SNR is 19 dB. By means of Eq. (21.31), a PIN receiver has a sensitivity of

$$\begin{aligned} P_r (\text{dB}) &= -118 \text{ dB}_m + 10 \log(36 \text{ MHz}) + 9.5 \text{ dB} - 10 \log 0.7 \\ &= -34.5 \text{ dB}_m = 0.36 \mu\text{W} \end{aligned}$$

For an APD receiver, the NEP is  $-132 \text{ dB}_m$ , which results in a sensitivity of

$$\begin{aligned} P_r &= -132 \text{ dB}_m + 10 \log(36 \text{ MHz}) + 9.5 \text{ dB} - 10 \log 0.7 \\ &= -48.4 \text{ dB}_m = 0.014 \mu\text{W} \end{aligned}$$

The receiver sensitivity sets the minimum allowed optical power incident on the photodetector and has a direct bearing on the selecting of the transmitter and fiber. The difference between the power coupled into the fiber and the sensitivity gives the maximum loss allowed between the transmitter and the receiver. Characteristics of photodetectors and receivers are summarized in Tables 21.5 to 21.7.

## 21.5 SELECTED BIBLIOGRAPHY

- Aggarwal, I. D.: "Recent Developments in Optical Fibers," *Laser Focus*, vol. 18, no. 11, 1982, pp. 115-117.
- Barnoski, M. K. (ed.): *Fundamentals of Optical Fiber Communications*, 2d ed., Academic, New York, 1981.
- Barnoski, M. K.: "Fiber Systems for the Military Environment," *Proceedings of the IEEE*, vol. 68, no. 10, 1980, pp. 1315-1320.
- Bliss, J., and D. W. Stevenson: "Fiber Optics: A Designer's Guide to System Budgeting," *Electro-Optical System Design*, vol. 13, no. 8, 1981, pp. 23-32.
- Campbell, L., and A. Barnhart: "Fiber Optic Transmission Link Analysis," *Proceedings of the SPIE*, vol. 224, pp. 12-18 (April 1980 meeting in Washington).
- Channin, D. J.: "Emitters for Fiberoptic Communications," *Laser Focus*, vol. 18, no. 11, 1982, pp. 105-113.
- CSELT Technical Staff, *Optical Fibre Communications*, McGraw-Hill, New York, 1980.
- Elion, G. R., and H. A. Elion: *Fiber Optics in Communications Systems*, Dekker, New York, 1978.
- Forrest, S. R.: "Photodiodes for Long-Wavelength Communication Systems," *Laser Focus*, vol. 18, no. 12, 1982, pp. 81-90.
- Giallorenzi, T. C.: "Fiber Optic Communications," *Proceedings of the IEEE*, vol. 66, no. 7, 1978, pp. 744-780.

- Hewlett-Packard Application Note 1000: "Digital Data Transmission with the HP Fiber Optic System," Hewlett-Packard Components, 640 Page Mill Road, Palo Alto, Calif. 94304.
- Howes, M. J., and D. V. Morgan (eds.): *Optical Fibre Communications: Devices, Circuits and Systems*, Wiley, New York, 1980.
- Kao, C. K.: *Optical Fiber Systems*, McGraw-Hill, New York, 1982.
- Lacy, E. A.: *Fiber Optics*, Prentice-Hall, Englewood Cliffs, N.J., 1982.
- Miller, S. E., and A. G. Chynoweth (eds.): *Optical Fiber Telecommunications*, Academic, New York, 1979.
- Personick, S. D.: *Optical Fiber Transmission Systems*, Plenum, New York, 1981.
- Personick, S. D., N. L. Rhodes, D. C. Hanson, and K. H. Chan, "Contrasting Fiber-Optic Component-Design Requirements in Telecommunications, Analogue and Local Data Communications Applications," *Proceedings of the IEEE*, vol. 68, no. 10, 1980, pp. 1254-1262.
- Seippel, R. G.: *Optoelectronics*, Reston Publishing, Reston, VA, 1981.
- Sharma, A. B., S. J. Halme, and M. M. Butusov: *Optical Fiber Systems and Their Components*, Springer-Verlag, New York, 1981.
- Tebo, A. R.: "Sensing with Optical Fibers: An Emerging Technology," *Electrooptical System Design*, vol. 14, no. 2, 1982, pp. 39-45.

Additional technical literature on fiber optics may be found in the following special journal issues:

- The Western Electric Engineer*, winter 1980.
- Proceedings of the IEEE*, vol. 68, no. 10, 1980.
- IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-30, no. 4, April 1982; no. 10, October 1982.
- IEEE Journal of Lightwave Technology*.





# Chapter 22

## Active Filters

**Roger C. Thielking**

*Professor of Electrical Technology  
Onondaga Community College  
Syracuse, New York*

### 22.1 INTRODUCTION

An active filter has circuitry that contains one or more active components. In most cases, operational amplifiers (op amps) are used. Compared with passive filters, active filters have these potential advantages:

1. Voltage amplification is often possible.
2. Output impedance is very low, and input impedance is usually very high. Thus the filter performance is unlikely to be affected by the signal source or load characteristics.
3. Inductors are not needed. Any desired characteristics may be obtained by using only resistors, capacitors, and op amps.
4. Capacitors are usually small, even for low frequency applications.

At high frequencies, however, performance is limited by the gain-bandwidth product of the op amps. For frequencies approaching 1 MHz and above, passive filters may be better.

Passive filters usually include both capacitors and inductors, which, being large, are costly if low frequencies are involved. With active networks, small values of capacitance may be used for the same functions, as illustrated with two basic circuits. Figure 22.1a shows a capacitance multiplier circuit. The input impedance  $Z_{in}$  is equivalent to a resistor  $R'$  in series with a capacitor  $C'$ , as shown in Fig. 22.1b. At low frequencies where  $R' < X_C$ , the resistance can be neglected and the circuit acts as a capacitor to ground, with the value of  $C_1$  multiplied by the ratio  $R_1/R_2$ . With the values shown, the multiplication factor is 1000, or  $10^3$ . A simulated inductor is illustrated in Fig. 22.2a. Op amp  $OA_1$  inverts the phase of the input current, which effectively shifts the phase angle of the input impedance by  $180^\circ$  so that it looks inductive rather than capacitive. The equivalent circuit is shown in Fig. 22.2b. Resistors  $R_1$  and  $R_2$  influence  $Q$  of the equivalent inductance  $L'$  as well as its value. For the values shown,  $L' = 10$  H.

There are a very large number of possible active filter circuits. Complete analysis and derivation of formulas are usually rather complex and difficult. In this chapter, selected filter configurations that are representative of various types, and among the most useful, are shown. Design data and formulas are given without derivation. Phase shift data, which are usually of less interest, are not generally shown.

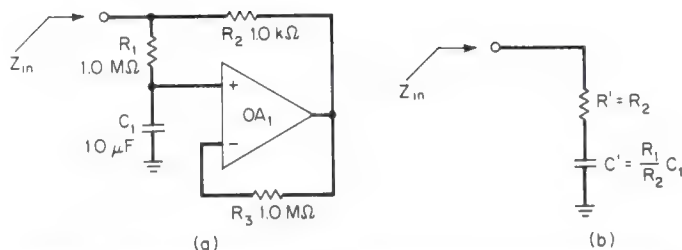


Fig. 22.1 Capacitance multiplier. (a) Circuit. (b) Equivalent input impedance.

## 22.2 LOW-PASS AND HIGH-PASS FILTERS

Ideally, a low-pass filter will pass all frequencies within its passband (below the cutoff frequency  $f_c$ ) without attenuation and will reject all higher frequencies. Conversely, in a high-pass filter, frequencies below  $f_c$  are rejected and all higher frequencies pass unchanged. (See Fig. 13.1a and b.) High-frequency response in active filters is, however, generally limited by the high-frequency response of the op amps used.

**FIRST-ORDER FILTERS** These are perhaps the simplest of active filters, consisting of a single *pole*, or RC (resistor-capacitor) combination, and an op amp connected as a voltage follower for low-output impedance, to isolate the load.

The amplitude and phase characteristics of a filter are given by its *transfer function*, the ratio of output voltage and phase to input voltage and phase. The general transfer functions for first-order low- and high-pass filters are as follows:

$$\text{Low-pass filter} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_l \omega_c}{s + \omega_c} \quad (22.1)$$

$$\text{High-pass filter} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_h s}{s + \omega_c} \quad (22.2)$$

The operator is  $s = j\omega$ , where  $j = \sqrt{-1}$  and  $\omega$  is frequency in radians per second;  $\omega_c$  is the cutoff radian frequency; and  $A_l$  and  $A_h$  are the passband voltage gains. Radian frequencies are related to frequencies in hertz by  $\omega = 2\pi f$  and  $\omega_c = 2\pi f_c$ .

A low-pass filter is shown in Fig. 22.3a. For this circuit, passband gain  $A_l = 1$ , the cutoff radian frequency is  $\omega_c = 1/(RC)$ , and the cutoff frequency in hertz is

$$f_c = \frac{1}{2\pi RC} \quad (22.3)$$

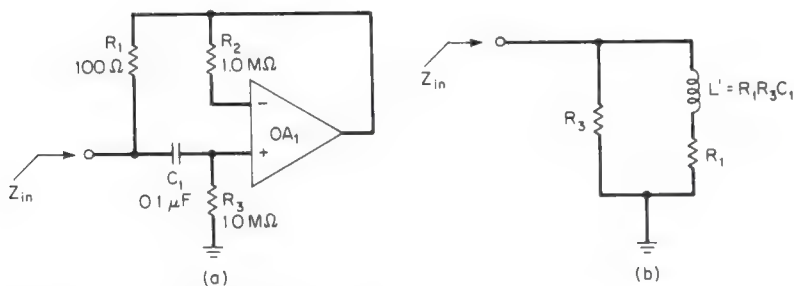
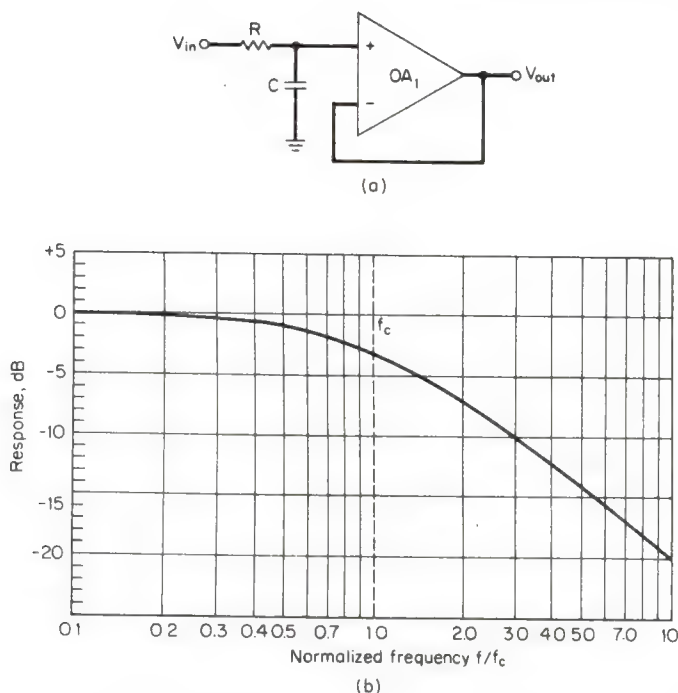


Fig. 22.2 Simulated inductance. (a) Circuit. (b) Equivalent input impedance.



**Fig. 22.3** First order low-pass filter with unity gain. (a) Circuit. (b) Frequency response, normalized to  $f_c = 1$  Hz.

Relatively high impedance values may be used for  $R$  and  $C$ , since the op amp  $OA_1$  does not appreciably load the network. Frequency response is determined by Eq. (22.1) and is shown in Fig. 22.3b, with frequency normalized to  $f_c$ .

Figure 22.4a shows the high-pass filter. The passband gain is  $A_h = 1$ , the cutoff radian frequency is  $\omega_c = 1/(RC)$ , and  $f_c$  is given by Eq. (22.3). The transfer function, Eq. (22.2), produces the frequency response of Fig. 22.4b.

**example 22.1** Design a low-pass first-order filter with a cutoff frequency  $f_c = 800$  Hz. If the input  $V_{in}$  is a sine wave with frequency  $f = 1.6$  kHz and 5.0 V peak-to-peak (p-p) amplitude, what is the amplitude of the output voltage  $V_{out}$ ?

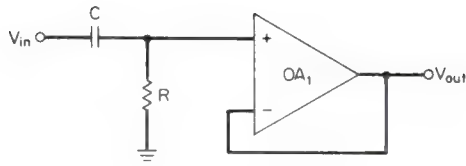
**solution** From Eq. (22.3),  $C = 1/(2\pi f_c R)$ . Select a reasonable value for  $R$ , say 1.0 M $\Omega$ . Substituting values, we find  $C = 1/(2\pi \cdot 800 \text{ Hz} \cdot 1.0 \text{ M}\Omega) = 199$  pF. Use these values in the circuit of Fig. 22.3a.

At the signal frequency,  $f/f_c = 1.6 \text{ kHz}/800 \text{ Hz} = 2.0$ . From Fig. 22.3b, the attenuation is approximately  $-6.7 \text{ dB} = 20 \log(V_{out}/5.0 \text{ V})$ . Solving yields  $V_{out} = 5.0 \text{ V}[\log^{-1}(-6.7 \text{ dB}/20)] \approx 2.3 \text{ V p-p}$ .

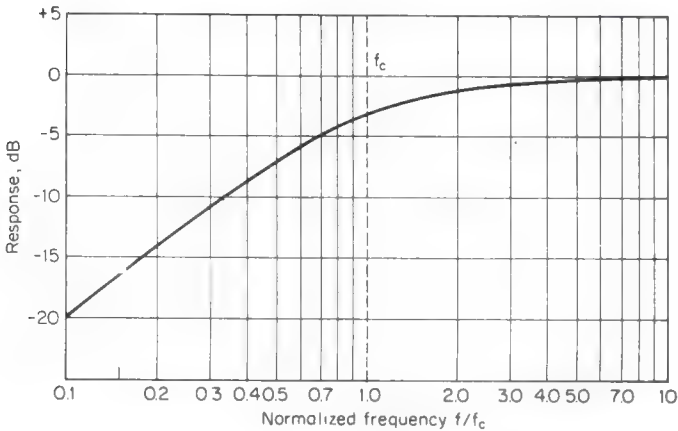
**example 22.2** A high-pass filter (Fig. 22.4a) has  $R = 220 \text{ k}\Omega$  and  $C = 10 \text{ nF}$ . Find the cutoff frequency  $f_c$  and the frequency where the gain is reduced to 20 percent of the passband value.

**solution** From Eq. (22.3), the cutoff frequency is  $f_c = 1/(2\pi \cdot 220 \text{ k}\Omega \cdot 10 \text{ nF}) = 72.3 \text{ Hz}$ . For 20 percent gain,  $-20 \log 0.2 = -14.0 \text{ dB}$ . From Fig. 22.4b,  $f/f_c = f/72.3 \text{ Hz} \approx 0.20$ . Solving gives  $f \approx 14.5 \text{ Hz}$ .

**Amplifying filters** First-order filters may be modified to provide voltage amplification by adding two resistors in the op amp feedback, as shown in Fig. 22.5. In the transfer functions, Eqs. (22.1) and (22.2), the gain factors  $A_l = A_h = (R_2 + R_3)/R_3$ . Equation (22.3) becomes  $f_c = 1/(2\pi R_1 C)$ . The frequency responses, normalized to passband gain, are as shown in Figs. 22.3b and 22.4b.

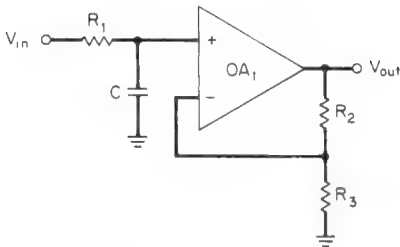


(a)

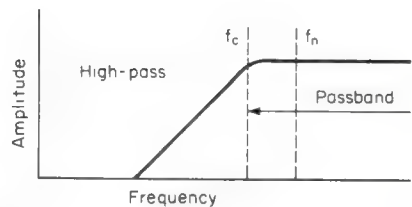
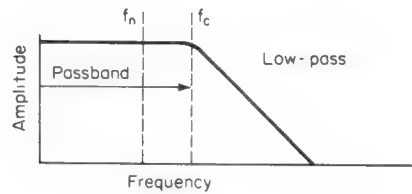


(b)

**Fig. 22.4** First-order high-pass filter with unity gain. (a) Circuit. (b) Frequency response, normalized to  $f_c = \text{Hz}$ .



**Fig. 22.5** First-order low-pass filter, modified for variable gain (for high-pass,  $R_1$  and  $C$  are interchanged).



**Fig. 22.6** Relationship of natural frequency  $f_n$  and cutoff frequency  $f_c$  to passband (typical).

**SECOND-ORDER FILTERS** Probably the most widely used circuit modules are those with two poles, referred to as *second-order* filters. Although only resistors and capacitors are used, the circuits are configured so that one of the poles acts like the inductive reactance of a second-order passive filter. The transfer functions are as follows:

$$\text{Low-pass} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_l \omega_1 \omega_2}{(s + \omega_1)(s + \omega_2)} \quad (22.4)$$

$$\text{High-pass} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_h s^2}{(s + \omega_1)(s + \omega_2)} \quad (22.5)$$

where  $\omega_1$  and  $\omega_2$  are the two pole frequencies (in radians per second). The *natural frequency*, or resonant frequency,  $\omega_n$  is given by

$$\omega_n = \sqrt{\omega_1 \omega_2} \quad (22.6)$$

the transfer functions also may be expressed as follows:

$$\text{Low-pass} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_l \omega_n^2}{s^2 + (\omega_n/Q)s + \omega_n^2} \quad (22.7)$$

$$\text{High-pass} \quad \frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_h s^2}{s^2 + (\omega_n/Q)s + \omega_n^2} \quad (22.8)$$

where  $Q$  is the *quality factor* of the circuit. Sometimes the *damping factor*  $\zeta$  is used instead. It is related to  $Q$  by  $\zeta = 1/(2Q)$ .

The natural frequency in hertz is  $f_n = \omega_n/(2\pi)$ . The passband may extend somewhat beyond  $f_n$  (see Fig. 22.6). Its edge corresponds to the *cutoff frequency*  $f_c$  (hertz) or  $\omega_c$  (radians per second). The natural and cutoff frequencies are related by the passband extension factor  $k_c$  as follows:

$$\text{Low-pass} \quad \omega_c = k_c \omega_n \quad \text{or} \quad f_c = k_c f_n \quad (22.9)$$

$$\text{High-pass} \quad \omega_c = \frac{\omega_n}{k_c} \quad \text{or} \quad f_c = \frac{f_n}{k_c} \quad (22.10)$$

Now  $k_c$  is related to  $Q$  as shown in Table 22.1 and depends on circuit values.

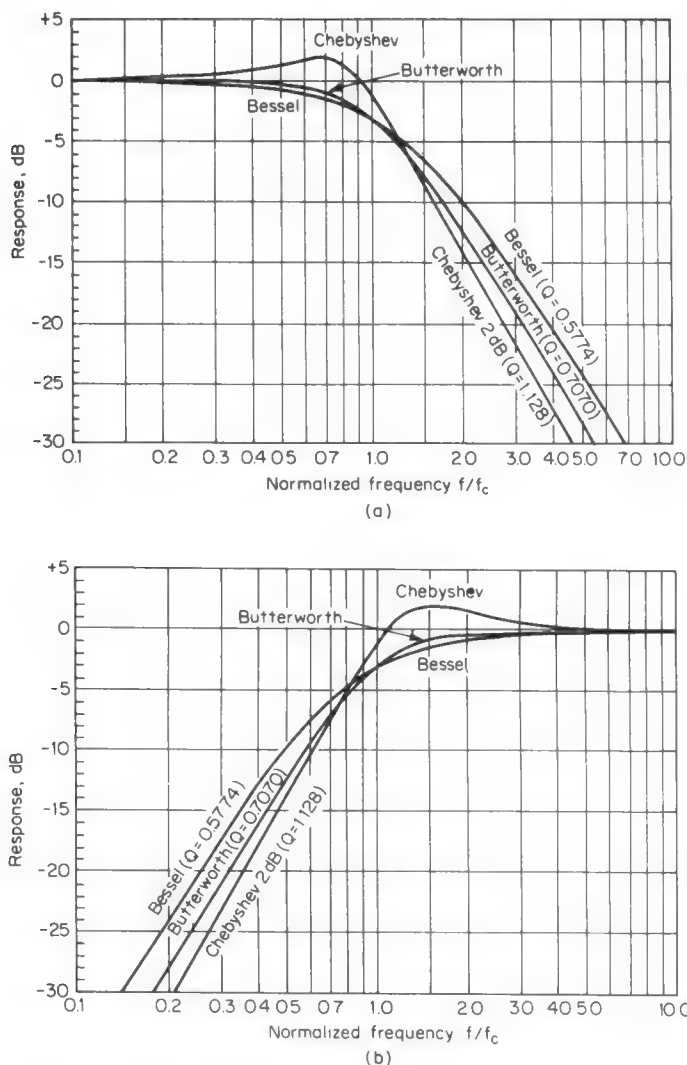
**Effect of  $Q$**  All second-order low- and high-pass filters have a flat response at frequencies well within the passband and attenuate at the rate of 40 dB/decade at frequencies well beyond cutoff. Near cutoff, however, the response varies with different values of the quality factor  $Q$ . Response types for selected  $Q$  values are listed in Table 22.1. Chebyshev characteristics, which include a range of  $Q$  values, have a peak in their frequency response near cutoff, where the passband amplitude is exceeded. The decibel values in Table 22.1 are for these peaks.

The Bessel characteristic has the most nearly constant signal delay for all passband frequencies. With the Butterworth characteristic, the amplitude response has maximum flatness within the passband. The Chebyshev characteristics have a more abrupt passband edge and slightly greater attenuation beyond, together with the response peak. Response curves for the Bessel, Butterworth, and 2-dB Chebyshev filters are shown in Fig. 22.7. These cor-

**TABLE 22.1 Filter Response Types**

Filter response type	Quality factor $Q$	Passband extension factor $k_c$
Bessel	0.5774	0.7851
Butterworth	0.7070	1.000
Chebyshev 0.5-dB peak	0.8637	1.128
Chebyshev 1.0-dB peak	0.9564	1.159
Chebyshev 2.0-dB peak	1.128	1.184

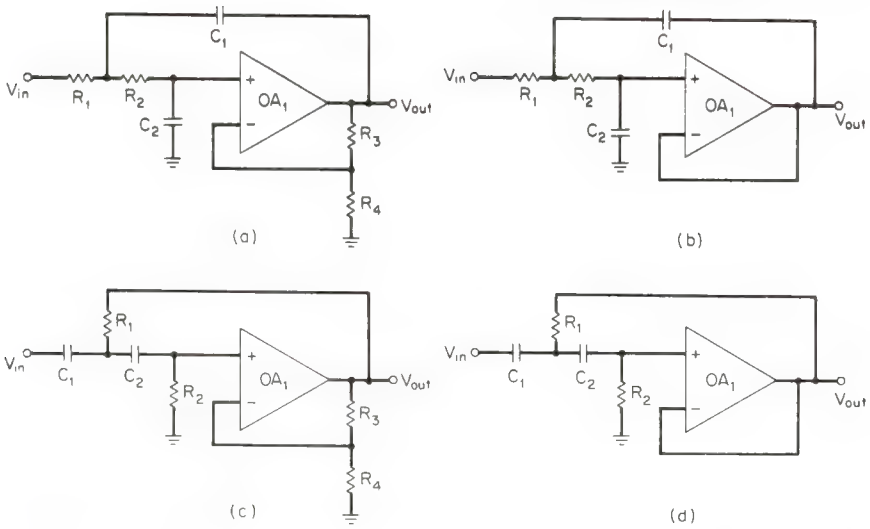




**Fig. 22.7** Response characteristics of second-order filters, normalized to  $f_c = 1$  Hz and unity gain. (a) Low-pass. (b) High-pass.

respond with filter  $Q$ 's of 0.5574, 0.7070, and 1.128, respectively. Filters with  $Q$  between these values would have intermediary response curves. These are "universal" curves, normalized to the cutoff frequency and passband amplitude, so that they are easily applied to any particular filter values. They apply to all the low- and high-pass circuits described below.

**Circuits** There are a number of possible circuits for second-order active filters. The *single-feedback* circuits, also called the *voltage-controlled voltage source* (VCVS) of Fig. 22.8 and the *infinite gain multiple feedback* (IGMF) circuits of Fig. 22.9, are representative, widely used types. (Another very important type is the *variable-state* filter.) These circuits may be analyzed by determining the natural frequency  $f_n$ , passband gain  $A_i$  or  $A_h$ , and



**Fig. 22.8** Single-feedback second-order filters. (a) Low-pass, variable-gain filter. (b) Low-pass, unity-gain filter. (c) High-pass, variable-gain filter. (d) High-pass, unity-gain filter.

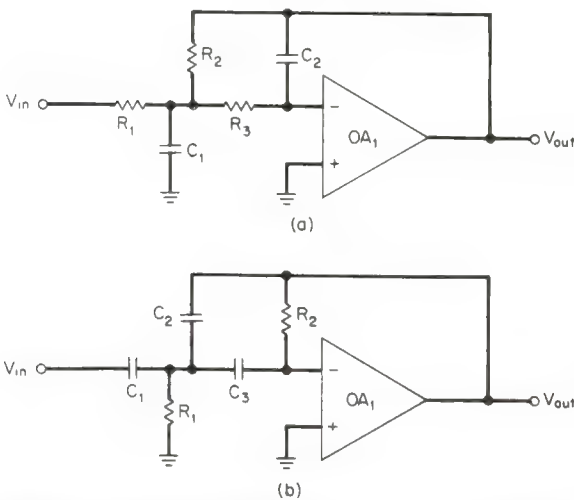
quality factor  $Q$ . Then, from Table 22.1 and Eq. (22.9) or (22.10), the cutoff frequency  $f_c$  is calculated or estimated.

Formulas for the circuits are as follows:

VCVS low-pass variable gain (Fig. 22.8a):

$$\omega_n = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}} \quad (22.11)$$

$$Q = \frac{1}{\omega_n [R_2 C_2 + R_1 C_2 - (A_1 - 1) R_1 C_1]} \quad (22.12)$$



**Fig. 22.9** IGMF second-order filters. (a) Low-pass. (b) High-pass.

$$A_I = \frac{R_3 + R_4}{R_4} \quad (22.13)$$

VCVS low-pass unity gain (Fig. 22.8b):  $A_I = 1$ , Eq. (22.11) applies and

$$Q = \frac{1}{\omega_n C_2 (R_1 + R_2)} \quad (22.14)$$

VCVS high-pass variable gain (Fig. 22.8c): Eqs. (22.11) and (22.13) apply, and

$$Q = \frac{1}{\omega_n [R_1 C_1 + R_1 C_2 - (A_h - 1) R_2 C_2]} \quad (22.15)$$

VCVS high-pass unity gain (Fig. 22.8d):  $A_h = 1$ , Eq. (22.11) applies, and

$$Q = \frac{1}{\omega_n R_1 (C_1 + C_2)} \quad (22.16)$$

IGMF low-pass (Fig. 22.9a):

$$\omega_n = \frac{1}{\sqrt{R_2 R_3 C_1 C_2}} \quad (22.17)$$

$$A_I = -\frac{R_2}{R_1} \quad (22.18)$$

$$Q = \frac{\omega_n C_1}{1/R_1 + 1/R_2 + 1/R_3} \quad (22.19)$$

IGMF high-pass (Fig. 22.9b):

$$\omega_n = \frac{1}{\sqrt{R_1 R_2 C_2 C_3}} \quad (22.20)$$

$$A_h = -\frac{C_1}{C_2} \quad (22.21)$$

$$Q = \frac{1}{\omega_n R_1 (C_1 + C_2 + C_3)} \quad (22.22)$$

**example 22.3** Find the characteristics  $f_c$ ,  $A_h$ , and  $Q$  of the filter shown in Fig. 22.10.

**solution** Comparing this circuit with Fig. 22.8 indicates that it is a high-pass, single-feedback, variable-gain filter. Equations (22.11), (22.13), and (22.15) apply:

$$\omega_n = \frac{1}{\sqrt{140 \text{ k}\Omega \cdot 140 \text{ k}\Omega \cdot 1.0 \text{ nF} \cdot 1.0 \text{ nF}}} = 7143 \text{ rad/s}$$

$$A_h = \frac{200 \text{ k}\Omega + 168 \text{ k}\Omega}{200 \text{ k}\Omega} = 1.840$$

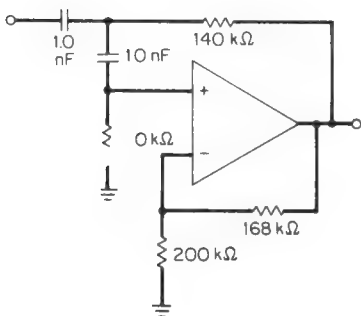


Fig. 22.10 Circuit for Example 22.3.

$$Q = \frac{1}{(7143 \text{ rad/s})[140 \text{ k}\Omega \cdot 1.0 \text{ nF} + 140 \text{ k}\Omega \cdot 1.0 \text{ nF} - (1.840 - 1)140 \text{ k}\Omega \cdot 1.0 \text{ nF}]}$$

$$= 0.8621$$

Comparison of this value of  $Q$  with the listing in Table 22.1 indicates a Chebyshev characteristic with about 0.5-dB peak and passband extension factor  $k_c \approx 1.128$ . From Eq. (22.10), the cutoff frequency is  $\omega_c \approx (7.143 \text{ krad/s})/1.128 \approx 6.332 \text{ krad/s}$ , corresponding to  $f_c \approx (6.332 \text{ krad/s})/2\pi \approx 1.0 \text{ kHz}$ .

**Design data** Filters may be designed by using the above equations for the various circuit types. The process is difficult, however, and may involve complex mathematical manipulations. Design is simplified with design aids, such as the data in Tables 22.2 through 22.5, applied to the circuits of Figs. 22.11, 22.12, and 22.13.

The design steps, by using the tables, are as follows:

1. Assume reasonable values for the components, as indicated on the referenced figure showing the schematic diagram.

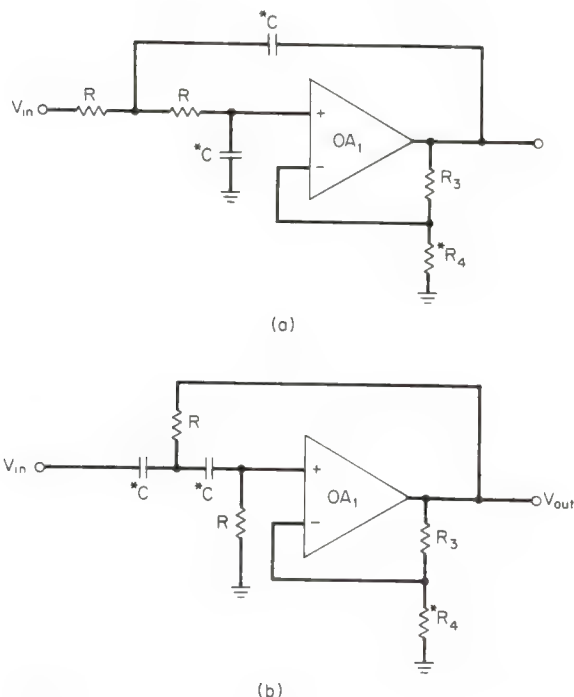
2. Calculate other component values by using the table. Primed symbols indicate *normalized* values (i.e., the  $R'$  value is in ohms when  $f_c = 1 \text{ Hz}$  and  $C = 1 \text{ F}$ ; to find  $R$ , divide  $R'$  by the actual values of  $f_c$  and  $C$ ).

3. If any of the resulting component values are too large or too small, they may be rescaled by multiplying resistance and dividing capacitance by the same factor.

**example 22.4** Using the circuit of Fig. 22.13a, design an IGMF low-pass filter with a Butterworth characteristic, cutoff frequency  $f_c = 2.5 \text{ kHz}$ , and gain  $A_1 = -2$ .

**solution** Select a value for  $C$ , say  $1.0 \mu\text{F}$ . Then from Table 22.4

$$R_1 = \frac{R'_1}{f_c C} = \frac{2.633}{2.5 \text{ kHz} \cdot 1.0 \mu\text{F}} = 1.05 \text{ k}\Omega$$



**Fig. 22.11** Single-feedback variable-gain circuits for design data (Table 22.2). (a) Low-pass. (b) High-pass. \* Assume values for  $C$  and  $R_4$ .

TABLE 22.2 Design Data for Single-Feedback Variable-Gain Filters

Characteristic	Either type (low-/high-pass)		Low-pass, Fig. 22.11a	High-pass, Fig. 22.11b
	$A_l/A_h$	$R_3$	$R'$	$R'$
Bessel	1.268	$0.2681R_4$	0.1250	0.2027
Butterworth	1.586	$0.5856R_4$	0.1592	0.1592
Chebyshev 0.5-dB peak	1.842	$0.8422R_4$	0.1795	0.1411
Chebyshev 1-dB peak	1.954	$0.9544R_4$	0.1845	0.1373
Chebyshev 2-dB	2.113	$1.113R_4$	0.1884	0.1344

Note:  $R = R'/f_c C$ .

TABLE 22.3 Design Data for Single-Feedback Unity-Gain Filters

Characteristic	Low-pass, Fig. 22.12a		High-pass, Fig. 22.12b	
	$C'_1$	$C'_2$	$R'_1$	$R'_2$
Bessel	0.1443	0.1082	0.1755	0.2341
Butterworth	0.2250	0.1126	0.1126	0.2250
Chebyshev 0.5-dB peak	0.3102	0.1040	0.08168	0.2437
Chebyshev 1-dB peak	0.3530	0.09646	0.07179	0.2627
Chebyshev 2-dB peak	0.4253	0.08349	0.05958	0.3033

Note:  $C = C'/f_c R$ ;  $R = R'/f_c C$ .

TABLE 22.4 Design Data for IGMF Low-Pass Filters, Fig. 22.13a

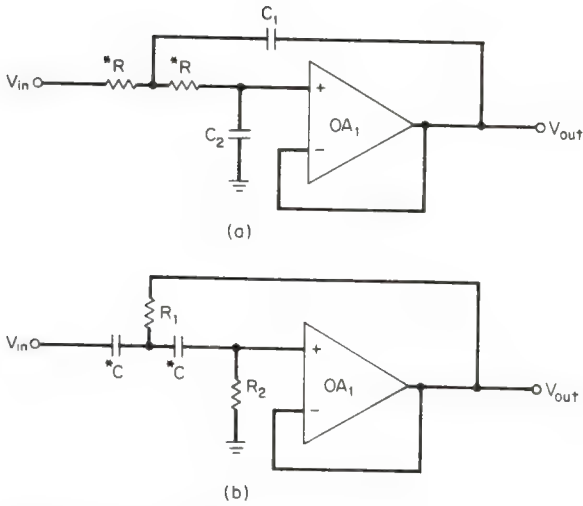
Characteristic	Unity gain, $A_l = -1$		$A_l = -2$	
	$R_2 = R_1$		$R_2 = 2R_1$	
	$R'_1$	$R'_2$	$R'_1$	$R'_2$
Bessel	5.263	0.07471	2.593	0.07528
Butterworth	5.393	0.1174	2.633	0.1203
Chebyshev 0.5-dB peak	4.865	0.1656	2.340	0.1722
Chebyshev 1-dB peak	4.438	0.1917	2.108	0.2017
Chebyshev 2-dB peak	3.695	0.2403	1.696	0.2618

Note:  $R = R'/f_c C$ .

TABLE 22.5 Design Data for IGMF High-Pass Filters, Fig. 22.13b

Characteristic	Unity gain, $A_h = -1$		$A_h = -2$	
	$C_1 = C$		$C_1 = 2C$	
	$R'_1$	$R'_2$	$R'_1$	$R'_2$
Bessel	0.1170	0.3512	0.08777	0.4682
Butterworth	0.07514	0.3376	0.05628	0.4501
Chebyshev 0.5-dB peak	0.05445	0.3656	0.04084	0.4875
Chebyshev 1-dB peak	0.04786	0.3940	0.03590	0.5253
Chebyshev 2-dB peak	0.03972	0.4549	0.02979	0.6065

Note:  $R = R'/f_c C$ .

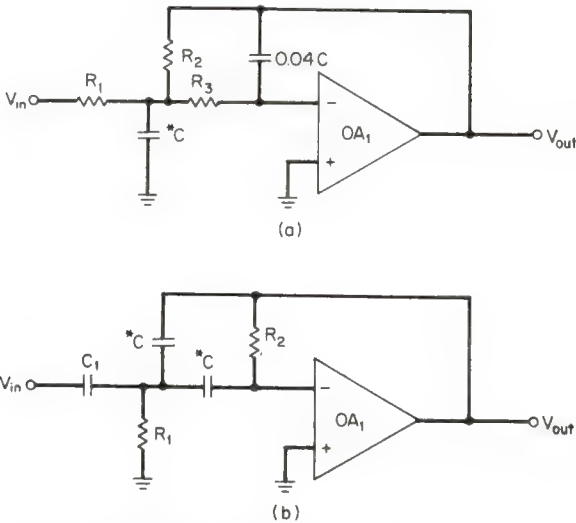


**Fig. 22.12** Single-feedback unity-gain circuits for design data (Table 22.3). (a) Low-pass (\*assume a value for  $R$ ). (b) High-pass (\*assume a value for  $C$ ).

$$R_3 = \frac{R'_3}{f_c C} = \frac{0.1203}{2.5 \text{ kHz} \cdot 1.0 \mu\text{F}} = 48.1 \Omega$$

$$R_2 = 2R_1 = 2 \cdot 1.05 \text{ k}\Omega = 2.10 \text{ k}\Omega$$

From Fig. 22.13a,  $C_1 = 0.04C = 40 \text{ nF}$ . These resistor values may be too low and cause excessive loading of the signal source. To rescale, multiply the resistance and divide the capacitance by 100. This yields  $C = 10 \text{ nF}$ ,  $C_1 = 400 \text{ pF}$ ,  $R_1 = 105 \text{ k}\Omega$ ,  $R_2 = 210 \text{ k}\Omega$ , and  $R_3 = 4.81 \text{ k}\Omega$ .



**Fig. 22.13** IGMF circuits for design data. (a) Low-pass (Table 22.4). (b) High-pass (Table 22.5). \*Assume a value for  $C$ .



**example 22.5** A noninverting unity-gain low-pass filter with a Chebyshev 1.0-dB-peak characteristic and a cutoff frequency of 50 Hz is needed. How much attenuation will it have at 125 Hz?

**solution** Use the single-feedback unity-gain low-pass filter circuit of Fig. 22.12a. Let  $R = 100 \text{ k}\Omega$ . Then, from Table 22.3,

$$C_1 = \frac{C'_1}{f_c R} = \frac{0.3530}{50 \text{ Hz} \cdot 100 \text{ k}\Omega} = 70.6 \text{ nF}$$

$$C_2 = \frac{C'_2}{f_c R} = \frac{0.09646}{50 \text{ Hz} \cdot 100 \text{ k}\Omega} = 19.3 \text{ nF}$$

At 125 Hz,  $f/f_c = 125 \text{ Hz}/50 \text{ Hz} = 2.5$ . Then, from Fig. 22.7a, attenuation is between  $-16 \text{ dB}$  (for Butterworth) and  $-18.3 \text{ dB}$  (for Chebyshev, 2-dB peak). A good estimate would be  $-17 \text{ dB}$ .

**Circuit features** In designing filters, the choice of circuits may be influenced by the following features:

Single-feedback (VCVS) unity-gain filter:

Noninverting

Fewest components

Single-feedback (VCVS) variable-gain filter:

Noninverting

Quality factor  $Q$  may be adjusted without changing the cutoff frequency

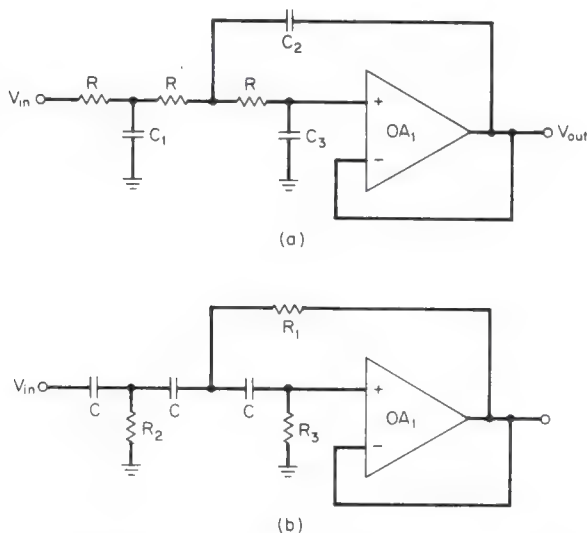
IGMF

Inverting

Fewest components for nonunity gain

**THIRD- AND HIGHER-ORDER FILTERS** Higher-order filters are more difficult to design and more expensive to construct, but they have better performance. As the order of the filter is increased, the sharpness of the passband edge increases, yielding more attenuation of the frequencies beyond.

**Single op amp third-order filter** Unity-gain low- and high-pass filters with three poles, or  $RC$  time constants, are shown in Fig. 22.14. Design data are shown in Tables 22.6 and 22.7. Frequency responses are plotted in Fig. 22.15.



**Fig. 22.14** Third-order filters. (a) Low-pass (for design data, see Table 22.6). (b) High-pass (for design data, see Table 22.7).

**TABLE 22.6 Design Data for Low-Pass Filter, Fig. 22.14a**

Characteristic	$C'_1$	$C'_2$	$C'_3$
Bessel	0.1572	0.1958	0.04039
Butterworth	0.2215	0.5644	0.03221
Chebyshev 0.5-dB peak	0.3581	1.787	0.01424
Chebyshev 1-dB peak	0.4086	2.575	0.01023
Chebyshev 2-dB peak	0.4954	4.428	0.006194

Note:  $C = C'/f_c R$ .

**example 22.6** For the high-pass third-order filter circuit shown in Fig. 22.14b, determine the component values needed for cutoff frequency  $f_c = 200$  Hz with a Chebyshev characteristic having a 1.0-dB peak.

**solution** Select a reasonable value for the capacitors, say  $C = 1.0$  nF. Then, from Table 22.7,

$$R_1 = \frac{0.009836}{1.0 \text{ nF} \cdot 200 \text{ Hz}} = 49.2 \text{ k}\Omega$$

$$R_2 = \frac{0.06201}{1.0 \text{ nF} \cdot 200 \text{ Hz}} = 310 \text{ k}\Omega$$

$$R_3 = \frac{2.476}{1.0 \text{ nF} \cdot 200 \text{ Hz}} = 12.4 \text{ M}\Omega$$

Resistor  $R_3$  is probably too large for a general-purpose op amp. Components may be rescaled by dividing resistances and multiplying capacitances by 20. Then  $C = 20$  nF,  $R_1 = 2.46$  k $\Omega$ ,  $R_2 = 15.5$  k $\Omega$ , and  $R_3 = 620$  k $\Omega$ .

Analysis of this type of circuit is difficult and is not considered here.

Fourth-order filters using a single op amp also may be designed, but the procedure is difficult and there are more possible problems with component values. The cascaded filter approach, described below, is usually preferred.

**Cascaded filters** A filter of any order may be devised by cascading first- and second-order filter circuits. For even-order filters, only second-order circuits are used, while one first-order circuit must be included to make an odd-order filter. Examples are shown in Fig. 22.16. Transfer functions, from Eqs. (22.1), (22.7), and (22.8), are indicated for each section. The overall transfer function is the product of the individual responses. Thus that for the third-order low-pass filter of Fig. 22.16a,

$$\frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_{11}A_{12}\omega_1^2\omega_2}{(s^2 + \omega_1s/Q_1 + \omega_1^2)(s + \omega_2)} \quad (22.23)$$

and for the sixth-order high-pass filter of Fig. 22.16b,

$$\frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_{h1}A_{h2}s^6}{(s^2 + \omega_1s/Q_1 + \omega_1^2)(s^2 + \omega_2s/Q_2 + \omega_2^2)(s^2 + \omega_3s/Q_3 + \omega_3^2)} \quad (22.24)$$

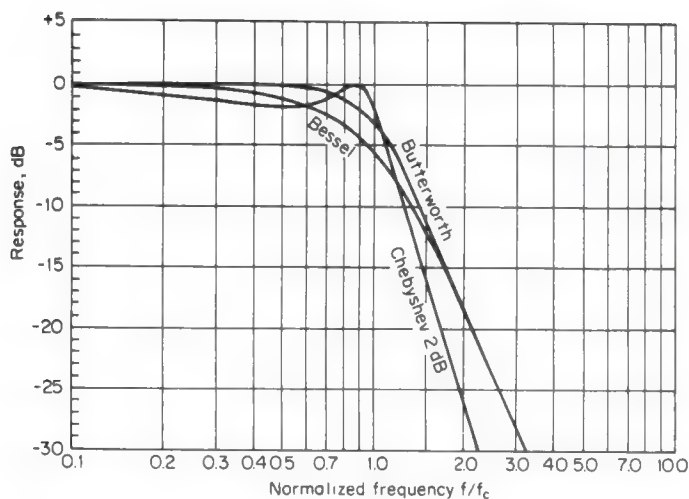
For filters of other orders, filter sections and transfer functions are similarly combined.

The  $Q$  values and pole frequencies, for various characteristics and orders of filters, are given in Table 22.8. These, together with the formulas for first- and second-order filters,

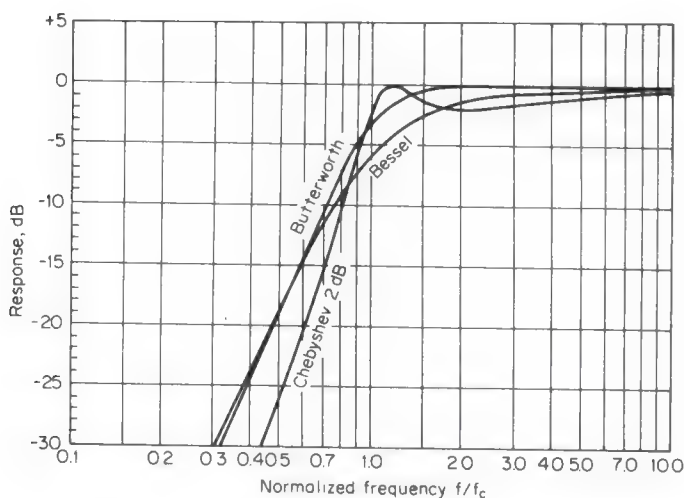
**TABLE 22.7 Design Data for High-Pass Filter, Fig. 22.14b**

Characteristic	$R'_1$	$R'_2$	$R'_3$
Bessel	0.1118	0.1611	0.6271
Butterworth	0.04488	0.1143	0.7864
Chebyshev 0.5-dB peak	0.01417	0.07073	1.778
Chebyshev 1-dB peak	0.009836	0.06201	2.476
Chebyshev 2-dB peak	0.005722	0.05112	4.089

Note:  $R = R'/f_c C$ .



(a)



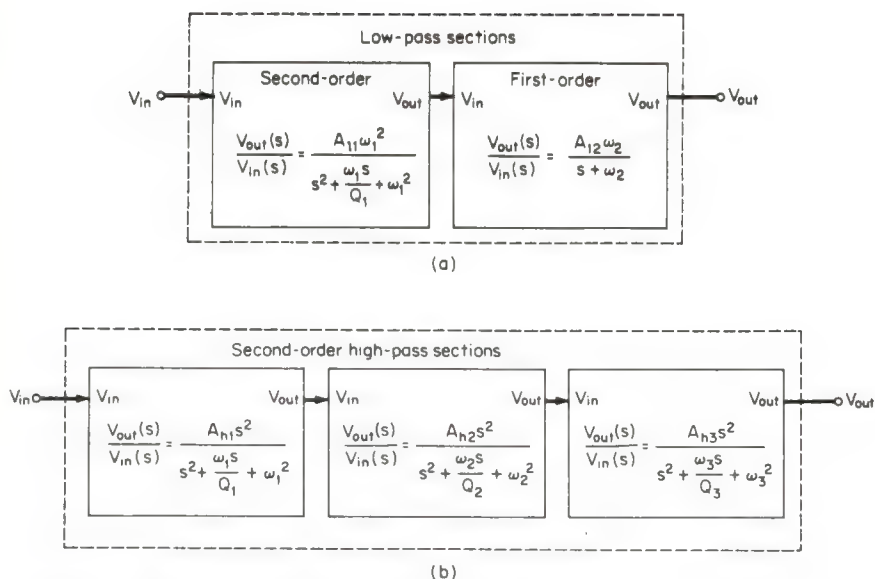
(b)

**Fig. 22.15** Response characteristics of third-order filters, normalized to  $f_c = 1$  Hz and unity gain. (a) Low-pass. (b) High-pass.

may be used for analysis and design. However, design with only the formulas is very difficult. Extensive tables and other design aids are available in the literature. An easier and perhaps better design technique is to use state-variable filters.

**Attenuation rate** As the order is increased, the attenuation rate of the response beyond cutoff also increases, as shown in Fig. 22.17. For the Butterworth characteristic, attenuation may be estimated with this formula:

$$\text{dB attenuation} = -20N \log \frac{f_1}{f_2} \quad (22.25)$$



**Fig. 22.16** Higher-order cascaded filters. (a) Third-order, low-pass. (b) Sixth-order, high-pass.

**TABLE 22.8** Design Data for Low-Pass Cascaded Filter Sections, Normalized to  $f_c = 1$  Hz

Order	Characteristic	$\omega'_1$	$Q_1$	$\omega'_2$	$Q_2$	$\omega'_3$	$Q_3$	$\omega'_4$	$Q_4$
2	Bessel	6.283	0.5774						
	Butterworth	6.283	0.7071						
	Chebyshev 0.5-dB peak	7.737	0.8637						
	Chebyshev 2-dB peak	5.700	1.129						
3	Bessel	6.778	0.6910	6.192					
	Butterworth	6.283	1.000	6.283					
	Chebyshev 0.5-dB peak	6.716	1.706	3.936					
	Chebyshev 2-dB peak	5.915	2.552	2.318					
4	Bessel	6.779	0.8055	6.046	0.9286				
	Butterworth	6.283	1.307	6.283	0.5412				
	Chebyshev 0.5-dB peak	6.480	2.941	3.751	0.7051				
	Chebyshev 2-dB peak	6.055	4.594	2.958	0.9294				
5	Bessel	6.818	0.9165	6.044	0.5635	5.835			
	Butterworth	6.283	1.618	6.283	0.6180	6.283			
	Chebyshev 0.5-dB peak	6.395	4.545	4.338	1.178	2.277			
	Chebyshev 2-dB peak	5.700	7.232	3.940	1.775	1.372			
6	Bessel	6.866	1.023	6.088	0.6112	5.781	0.5103		
	Butterworth	6.283	1.932	6.283	0.7071	6.283	0.5176		
	Chebyshev 0.5-dB peak	6.355	6.513	4.826	1.810	2.490	0.6836		
	Chebyshev 2-dB peak	6.175	10.46	4.587	2.844	1.986	0.9016		

continued

TABLE 22.8 Design Data for Low-Pass Cascaded Filter Sections, Normalized to  $f_c = 1$  Hz (Cont.)

Order	Characteristic	$\omega'_1$	$Q_1$	$\omega'_2$	$Q_2$	$\omega'_3$	$Q_3$	$\omega'_4$	$Q_4$
7	Bessel	6.914	1.126	6.148	0.6608	5.790	0.5324	5.682	
	Butterworth	6.283	2.247	6.283	0.8019	6.283	0.5550	6.283	
	Chebyshev 0.5-dB peak	6.334	8.842	5.169	2.576	3.166	1.092	1.610	
	Chebyshev 2-dB peak	6.203	14.28	5.008	4.115	2.896	1.646	0.9760	
8	Bessel	6.914	1.226	6.170	0.7019	5.788	0.5596	5.618	0.5060
	Butterworth	6.283	2.563	6.283	0.9000	6.283	0.6013	6.283	0.5098
	Chebyshev 0.5-dB peak	6.321	11.53	5.410	3.466	3.763	1.611	1.864	0.6766
	Chebyshev 2-dB peak	6.221	18.69	5.293	5.584	3.594	2.532	1.494	0.8924

Note:  $\omega_n(\text{low-pass}) = \omega' f_c$ ; for high-pass filter sections,  $\omega_n(\text{high-pass}) = 4\pi^2 f_c / \omega'$ .

where  $f_c = f_2$  (low-pass) or  $f_1$  (high-pass) and  $N$  is the order. There is an error of 3 dB maximum at  $f_c$ , which becomes negligible an octave or more beyond cutoff. The attenuation is less for Bessel and greater for Chebyshev characteristics (see Figs. 22.7 and 22.15) and increases similarly at higher orders.

**example 22.7** A Butterworth high-pass filter, with cutoff frequency  $f_c = 4.0$  kHz, must attenuate by at least  $-20$  dB at 2.2 kHz. Find the order necessary.

**solution** From Eq. (22.25),  $-20$  dB =  $-20N \log (4.0 \text{ kHz}/2.2 \text{ kHz}) = -20N(0.260)$ . Thus  $N = 3.85$ , and a fourth-order filter will suffice. (Although this is an estimate, any error would result in greater attenuation.)

22.3 BANDPASS AND BANDSTOP FILTERS

Band filters are characterized as *broadband* or *narrowband*, according to the width of the frequency band that is rejected or passed. Each requires different circuitry. The bandwidth is from  $f_{c1}$  to  $f_{c2}$ , the 3-dB cutoff frequencies. If the ratio  $f_{c2}/f_{c1} < 1.5$ , narrowband filters usually are best; if greater, broadband types may be used.

**BROADBAND FILTERS** For wide-frequency bands ( $f_{c2}/f_{c1} > 1.5$ ), filters are formed by combining the high- and low-pass filters described in Sec. 22.2. For bandpass, shown in Fig. 22.18, the low- and high-pass filters are cascaded. The low-pass cutoff frequency is  $f_{c2}$ , and the high-pass cutoff is  $f_{c1}$ . For bandstop or band-reject filters, the filters are connected in shunt as shown in Fig. 22.19. Op amp  $OA_1$  is used to combine the two outputs in a summing circuit. Here, the cutoff of the low-pass filter is the lower cutoff  $f_{c1}$  of the bandstop function.

In each case, the *skirt* of the response (the region of attenuation near cutoff) is determined by the response of its associated low- or high-pass filter. The other filter has very little effect,

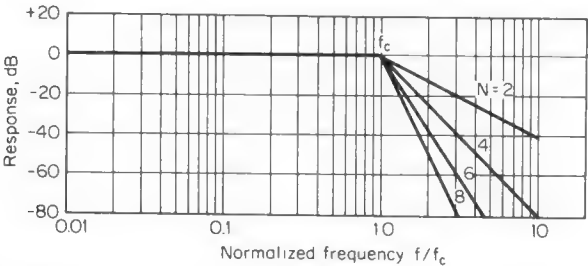


Fig. 22.17 Attenuation rates of passband skirt, low-pass Butterworth characteristic, for various orders.

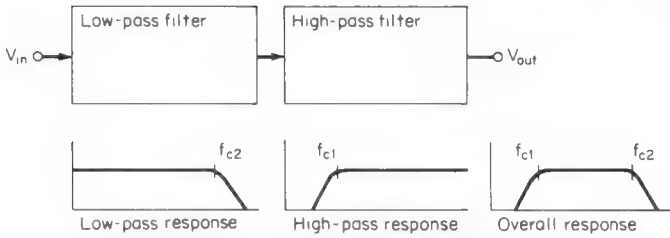


Fig. 22.18 Broadband bandpass filter.

except near the center frequency of the band-reject response. At this point, the filter outputs are about equal and are added, which can increase response (or reduce attenuation) by up to 6 dB. This effect can be reversed, so as to increase attenuation at the center frequency, with circuitry that causes phase reversal of the signals. Using odd-order filters will do this, or if an even-order filter is used, one filter can be of an inverting type while the other is noninverting.

**NARROWBAND FILTERS** When the cutoff frequencies are close together ( $f_{c2}/f_{c1} < 1.5$ ), a specially designed filter, which has a response similar to that of a tuned  $LC$  circuit, is usually used.

**Bandpass filter** The usual transfer function for a narrow-bandpass filter is

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{A_p s \omega_n}{s^2 + s \omega_n / Q + \omega_n^2} \quad (22.26)$$

Normalized frequency response is shown in Fig. 22.20 for various values of  $Q$ . These equations are useful:

Bandwidth:

$$f_{c2} - f_{c1} = \frac{f_n}{Q} \quad (22.27)$$

Natural resonant frequency:

$$f_n = \sqrt{f_{c1} f_{c2}} \quad (22.28)$$

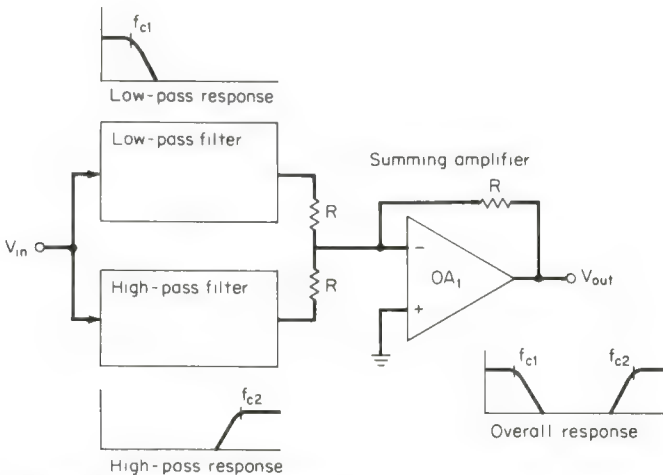
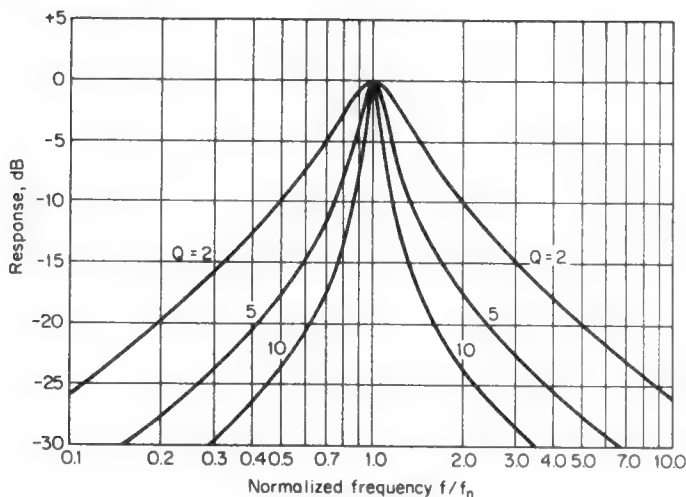


Fig. 22.19 Broadband bandstop filter.





**Fig. 22.20** Response characteristics, narrowband-bandpass filter, for various  $Q$  values.

Cutoff frequencies:

$$f_{c1} = f_n \frac{\sqrt{1 + 4Q^2} - 1}{2Q} \quad (22.29)$$

$$f_{c2} = f_n \frac{\sqrt{1 + 4Q^2} + 1}{2Q} \quad (22.30)$$

The attenuation in dB at any frequency  $f$  is

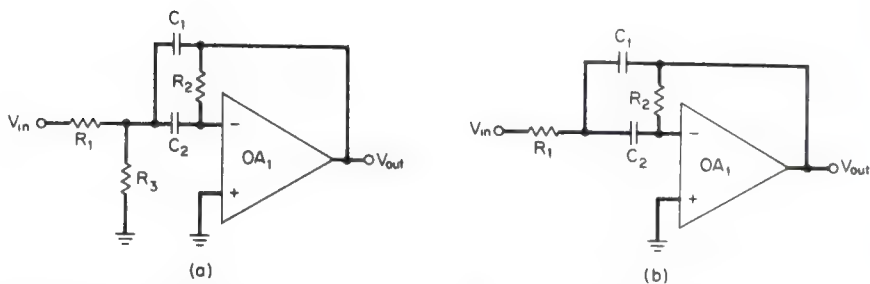
$$\text{dB attenuation} = -10 \log \left[ 1 + \frac{Q^2(f^2 - f_n^2)^2}{f^2 f_n^2} \right] \quad (22.31)$$

Two versions of a widely used circuit are given in Fig. 22.21. For the three-resistor version, Fig. 22.21a, the equations for analysis are

$$\omega_n = \left( \frac{R_1 + R_3}{C_1 C_2 R_1 R_2 R_3} \right)^{1/2} \quad (22.32)$$

$$A_p = \frac{C_2 R_2}{R_1 (C_1 + C_2)} \quad (22.33)$$

$$Q = \left[ \frac{R_1 + R_3}{R_1 R_2} \frac{R_2 C_1 C_2}{(C_1 + C_2)^2} \right]^{1/2} \quad (22.34)$$



**Fig. 22.21** Narrowband-bandpass filters. (a) Three-resistor circuit. (b) Two-resistor circuit.

For the two-resistor circuit of Fig. 22.21*b*, Eq. (22.33) applies. The others become

$$\omega_n = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}} \quad (22.35)$$

$$Q = \left[ \frac{R_2 C_1 C_2}{R_1 (C_1 + C_2)^2} \right]^{1/2} \quad (22.36)$$

**example 22.8** A bandpass filter with the circuit configuration of Fig. 22.21*b* has  $R_1 = 20 \text{ k}\Omega$ ,  $R_2 = 1.0 \text{ M}\Omega$ , and  $C_1 = C_2 = 3.5 \text{ nF}$ . Find the bandpass gain  $A_p$ , natural resonant frequency  $f_n$ , quality factor  $Q$ , bandwidth, cutoff frequencies  $f_{c1}$  and  $f_{c2}$ , and the attenuation at 1.0 kHz (relative to the maximum output at  $f_n$ ).

**solution** Equations (22.33), (22.35) and (22.36) apply. Substituting values, we find

$$\begin{aligned} A_p &= -\frac{1.0 \text{ M}\Omega \cdot 3.5 \text{ nF}}{20 \text{ k}\Omega(3.5 \text{ nF} + 3.5 \text{ nF})} = -25 \\ \omega_n &= \frac{1}{\sqrt{20 \text{ k}\Omega \cdot 1.0 \text{ M}\Omega \cdot 3.5 \text{ nF} \cdot 3.5 \text{ nF}}} = 2.02 \text{ krad/s} \\ f_n &= \frac{2.02 \text{ krad/s}}{2\pi} = 322 \text{ Hz} \\ Q &= \left[ \frac{1.0 \text{ M}\Omega \cdot 3.5 \text{ nF} \cdot 3.5 \text{ nF}}{20 \text{ k}\Omega(3.5 \text{ nF} + 3.5 \text{ nF})^2} \right]^{1/2} = 3.54 \end{aligned}$$

From Eqs. (22.27), (22.29), (22.30), and (22.31), we know that bandwidth =  $322 \text{ Hz}/3.54 = 90.8 \text{ Hz}$  and the cutoff frequencies are

$$f_{c1} = 322 \text{ Hz} \left( \frac{\sqrt{1 + 4 \cdot 3.54^2} - 1}{2 \cdot 3.54} \right) = 280 \text{ Hz}$$

$$f_{c2} = 322 \text{ Hz} \left( \frac{\sqrt{1 + 4 \cdot 3.54^2} + 1}{2 \cdot 3.54} \right) = 371 \text{ Hz}$$

$$\begin{aligned} \text{Attenuation at 1.0 kHz} &= -10 \log \left\{ 1 + 3.54^2 \frac{[(1.0 \text{ kHz})^2 - (322 \text{ Hz})^2]^2}{(322 \text{ Hz})^2 \cdot (1.0 \text{ kHz})^2} \right\} \text{ dB} \\ &= -19.9 \text{ dB} \end{aligned}$$

For design convenience, let  $C_1 = C_2 = C$ . The circuit of Fig. 22.21*a* is more versatile, allowing independent setting of gain  $A_p$  and quality factor  $Q$ . Design steps and formulas are as follows:

1. Assume a reasonable value for  $R_2$ , usually  $100 \text{ k}\Omega$  to  $1 \text{ M}\Omega$ .
2. Calculate  $R_1$  for the desired gain  $A_p$ :

$$R_1 = \frac{R_2}{2A_p} \quad (22.37)$$

3. Calculate  $R_3$  for the desired quality factor  $Q$ :

$$R_3 = \frac{R_1 R_2}{4R_1 Q^2 - R_2} \quad (22.38)$$

4. Calculate capacitance  $C$  for the desired natural resonant frequency  $f_n$ :

$$C = \frac{1}{2\pi f_n} \left( \frac{R_1 + R_3}{R_1 R_2 R_3} \right)^{1/2} \quad (22.39)$$

The circuit of Fig. 22.21*b* has fewer components, but  $Q$  and  $A_p$  are interdependent. Design steps and formulas are as follows:

1. Select a value for  $R_2$ , as above.
2. Calculate  $R_1$  for the desired quality factor  $Q$ :

$$R_1 = \frac{R_2}{4Q^2} \quad (22.40)$$

Alternatively,  $R_1$  may be calculated for a desired gain  $A_p$  with Eq. (22.37). In either case,  $A_p$  and  $Q$  are related by

$$A_p = 2Q^2 \quad (22.41)$$

3. Calculate  $C$  for the desired natural resonant frequency:

$$C = \frac{1}{2\pi f_n \sqrt{R_1 R_2}} \quad (22.42)$$

**example 22.9** Design a narrow-bandpass filter with unity gain (at  $f_n$ ) and with cutoff frequencies of 900 Hz and 1.11 kHz.

**solution** From Eq. (22.28),  $f_n = \sqrt{900 \text{ Hz} \cdot 1.11 \text{ kHz}} = 1.0 \text{ kHz}$ . From Eq. (22.27),  $Q = f_n/(f_{c2} - f_{c1}) = 1.0 \text{ kHz}/(1.11 \text{ kHz} - 900 \text{ Hz}) = 4.76$ . Equation (22.42) is not satisfied; therefore, the circuit of Fig. 22.21a must be used. We apply the design steps:

1. Assume  $R_2 = 1.0 \text{ M}\Omega$ .
2. From Eq. (22.37),  $R_1 = \frac{1}{2} \cdot 1.0 \text{ M}\Omega/1 = 500 \text{ k}\Omega$ .
3. From Eq. (22.38),

$$R_3 = \frac{1.0 \text{ M}\Omega \cdot 500 \text{ k}\Omega}{4 \cdot 500 \text{ k}\Omega \cdot 4.76^2 - 1.0 \text{ M}\Omega} = 11.3 \text{ k}\Omega$$

4. From Eq. (22.39),

$$C = \frac{1}{2\pi \cdot 1.0 \text{ kHz}} \left( \frac{500 \text{ k}\Omega + 11.3 \text{ k}\Omega}{500 \text{ k}\Omega \cdot 1.0 \text{ M}\Omega \cdot 11.3 \text{ k}\Omega} \right)^{1/2} = 1.51 \text{ nF}$$

These circuits work well for  $Q$  values up to 10, with 1 percent resistors and 2 percent capacitors.

**Bandstop filter** These are also called *band-elimination*, or *notch*, filters. The usual transfer function is

$$\frac{V_{\text{out}}(s)}{V_{\text{in}}(s)} = \frac{A_s(s^2 + \omega_n^2)}{s^2 + s\omega_n/Q + \omega_n^2} \quad (22.43)$$

Bandwidth and cutoff frequency considerations are the same as for the bandpass filter, so that Eqs. (22.27), (22.28), (22.29), and (22.30) apply. The attenuation is different. The normalized attenuation at any frequency  $f$  is

$$\text{dB attenuation} = -10 \log \left[ 1 + \frac{f^2 f_n^2}{Q^2 (f^2 - f_n^2)^2} \right] \quad (22.44)$$

The normalized frequency response is shown in Fig. 22.22 for various values of  $Q$ . Equation (22.44) indicates infinite attenuation at  $f_n$ . However, the practical maximum is  $-30$  to  $-40$  dB, depending on component precision.

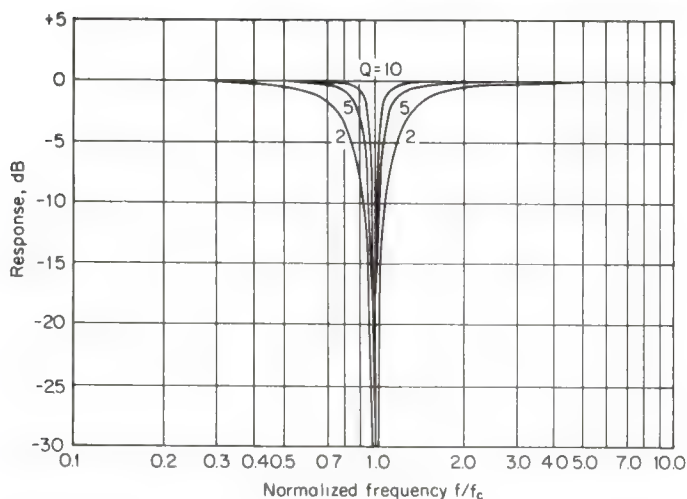
The bandpass filters of Fig. 22.21 may be used, with an op amp summing amplifier, to form a bandstop filter as shown in Fig. 22.23. The bandpass response must subtract from the flat response, with  $R_2 = |A_p|R_1$  ( $A_p$  is the passband gain of the bandpass filter). The bandstop characteristic has the same natural resonant frequency  $f_n$  and quality factor  $Q$  as the bandpass filter. The passband gain, of the overall bandstop is  $A_s = R_2/R_1$  and may be adjusted independently to any desired value.

Another circuit offering high performance is the *twin-T* configuration of Fig. 22.24. The single op amp version, Fig. 22.24a has a very low quality factor  $Q$  which is difficult to calculate, usually about 0.3 or somewhat less. In the more complex circuit of Fig. 22.24b,  $Q$  is increased by the bootstrapping function of op amp  $OA_2$ . Quality factor  $Q$  may be estimated by

$$Q \approx \frac{0.3(R_4 + R_5)}{R_4} \quad (22.45)$$

and has a practical upper limit of about 50, with precision components.

Both filter versions have unity passband gain, that is,  $A_s = 1$ . The natural resonant frequency, or notch frequency,  $f_n$  is



**Fig. 22.22** Response characteristics, narrowband bandstop filter, for various  $Q$  values.

$$f_n = \frac{1}{2\pi} \left( \frac{C_1 + C_2}{R_1 R_2 C_1 C_2 C_3} \right)^{1/2} \quad (22.46)$$

For design convenience with either circuit, use  $C_1 = C_2 = C_3 = C$ ,  $R_1 = R_2 = R$ , and  $R_3 = R/4$ . The design steps and formulas are as follows:

1. Select a value for  $R$ . Calculate  $C$ :

$$C = \frac{1}{\sqrt{2\pi R f_n}} \quad (22.47)$$

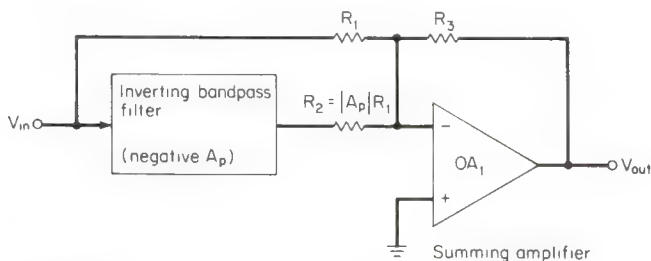
(The  $R$  and  $C$  values may be rescaled by maintaining the same  $RC$  product.)

2. The quality factor  $Q$  of the simpler circuit, Fig. 22.24a, is 0.238. This should be used if it is suitable.

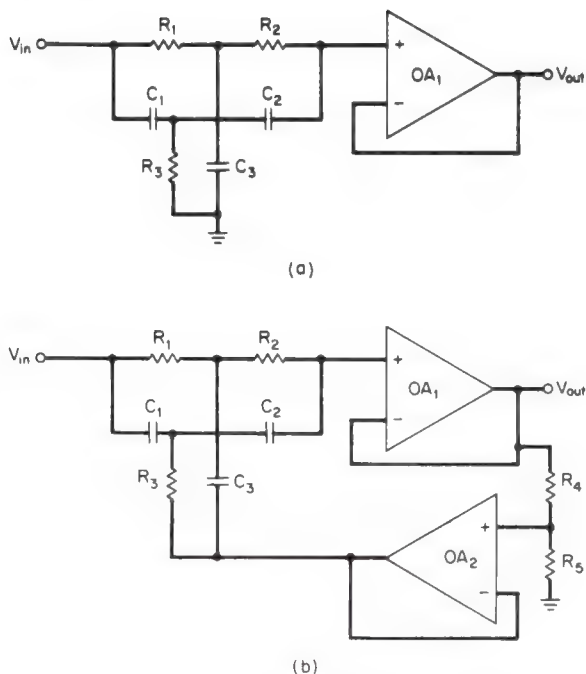
3. For higher  $Q$ , use the circuit of Fig. 22.24b. Select a value for  $R_5$ . Calculate  $R_4$  for the desired  $Q$ :

$$R_4 = \frac{\sqrt{2}R_5}{6Q - \sqrt{2}} \quad (22.48)$$

**example 22.10** Design a twin-T notch filter to have a notch frequency  $f_n = 15$  Hz and bandwidth = 1.3 Hz. Find its attenuation at 13 Hz.



**Fig. 22.23** Narrowband bandstop filter using a bandpass filter.



**Fig. 22.24** Twin-T bandstop filter. (a) Low- $Q$  circuit. (b) Bootstrapped circuit.

**solution** From Eq. (22.27),  $Q = f_n/\text{bandwidth} = 15 \text{ Hz}/1.3 \text{ Hz} = 11.5$ . Select  $R_1 = R_2 = R = 400 \text{ k}\Omega$  and  $R_3 = R/4 = 100 \text{ k}\Omega$ . From Eq. (22.47),  $C_1 = C_2 = C_3 = C = 1/(\sqrt{2}\pi \cdot 400 \text{ k}\Omega \cdot 15 \text{ Hz}) = 37.5 \text{ nF}$ .

Choose  $R_5 = 10 \text{ k}\Omega$ . Then, from Eq. (22.48),  $R_4 = (\sqrt{2} \cdot 10 \text{ k}\Omega)/(6 \cdot 11.5 - \sqrt{2}) = 209 \Omega$ . The attenuation at 13 Hz, from Eq. (22.44), is

$$\text{dB} = -10 \log \left\{ 1 + \frac{(13 \text{ Hz})^2 (15 \text{ Hz})^2}{(11.5)^2 [(13 \text{ Hz})^2 - (15 \text{ Hz})^2]^2} \right\} = -0.38 \text{ dB}$$

Narrowband filters may also be implemented with state-variable circuits.

## 22.4 REFERENCES AND BIBLIOGRAPHY

- Berlin, H. M.: *Design of Active Filters, with Experiments*, Sams, Indianapolis, 1978.
- Daniels, R. W.: *Approximation Methods for Electronic Filter Design*, Artech House, Dedham, Mass., 1975.
- Daryanani, G.: *Principles of Active Network Synthesis and Design*, Wiley, New York, 1976.
- Johnson, D. E.: *A Handbook of Active Filters*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
- Johnson, D. E., and J. L. Hilburn: *Manual of Active Filter Design*, McGraw-Hill, New York, 1973.
- Johnson, D. E., and J. L. Hilburn: *Rapid Practical Design of Active Filters*, Wiley, New York, 1975.
- Schaumann, R.: *Modern Active Filter Design*, Wiley, New York, 1981.
- Schaumann, R.: *Modern Active Filter Design*, Institute of Electrical and Electronics Engineers, New York, 1981.
- Williams, A. B.: *Active Filter Design*, Artech House, Dedham, Mass., 1975.

# Chapter 23

## Digital Multimeters

**Kenneth Jessen**

*Material Engineer  
Hewlett-Packard Co.  
Loveland, Colorado*

### 23.1 BASIC BUILDING BLOCKS

The job of a digital multimeter (DMM) is to convert an analog signal to its digital equivalent. The analog signal might be a dc voltage, ac voltage, or ac or dc current or resistance. The process of conversion can be divided into four functional blocks, as shown in Fig. 23.1.

The input signal first passes through some type of signal conditioner or converter. For dc voltages, the signal conditioner is composed of an attenuator, operational amplifier (op amp), and precision feedback resistors. The attenuator is used to divide higher voltages above the range of the op amp while lower voltages are amplified. Ranges in a DMM are usually in decade increments. If the input is an ac signal, an ac-to-dc (A/D) converter is used to convert the ac signal to its equivalent dc value. To measure resistance, an ohm converter supplies a constant current to the unknown resistor. Then the voltage drop across the resistor is measured as a dc voltage. In all cases, it is the job of the signal conditioner or converter to normalize the input to the DMM into a dc voltage within the range of the A/D converter.

The A/D converter takes the prescaled dc voltage and converts it to digits. These single-range devices operate at a full-scale voltage such as 1, 3, or 10 V depending on the type of DMM.

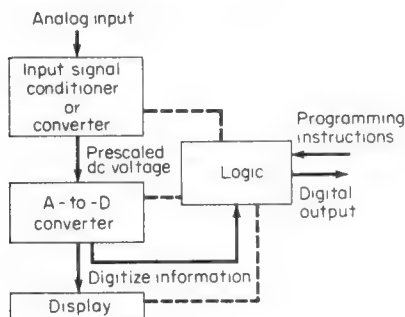
At this point it is wise to take a brief look at the mechanics of the conversion process. Suppose a 250-V ac signal is applied to the DMM input which has an A/D converter with a 1-V full-scale input. First the ac signal must be attenuated on the 1000-V ac range to 0.25 V ac (100:1 division ratio). Next the ac signal is converted to its dc equivalent. The op amp with its gain set to unity passes the dc signal to the A/D converter which digitizes the voltage to 2500. On the DMM display, range and function information is added to produce a display of 250.0 V ac.

If the signal happened to be a dc voltage of say 0.05 V, the op amp would be set to a gain of X10 such that 0.5 V dc would be passed on to the A/D converter. The digitized value would be 500, and again range and function information would be added by the display to produce 0.05 V dc, or 50 mV dc.

The first two building blocks govern most of the characteristics of a DMM such as accuracy, number of digits, ranges, sensitivity, and so on. The logic manages the flow of information to and from the instrument as well as internal functions. The job of the display is to visually communicate the result of a given measurement.



## 23.2 Digital Multimeters



**Fig. 23.1** A digital multimeter may be divided into four basic blocks as shown here. The input signal conditioner or converter normalizes the input to a dc voltage within the range of the A/D converter. The digitized output from the A/D converter is sent to the display as a visual output, and the logic manages the internal timing and flow of information into and out of the instrument.

**example 23.1** Show how a 500-V ac signal would be handled by a DMM with an A/D converter which had a 10-V full-scale input.

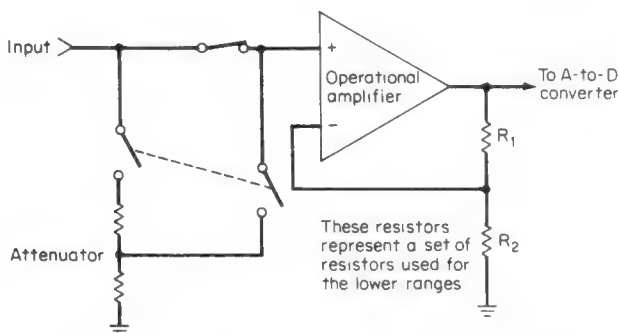
**solution** The 500 V ac would be attenuated on the 1000-V ac range to 5 V and converted to its dc equivalent. The op amp with its gain set to unity would pass this signal to the A/D converter. The signal would be digitized to 5000 and displayed as 500.0 V ac, using range and function information.

### 23.2 DC INPUT AMPLIFIER/ATTENUATOR

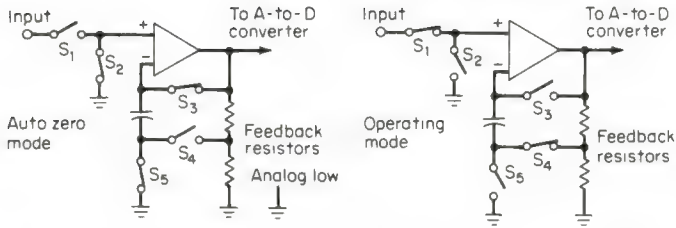
The signal conditioning for dc voltages includes both amplification and attenuation. If the full-scale input to the A/D converter is 10 V, then amplification is used on the lower ranges such as 100 mV and 1 V. Unity gain is used on the 10-V range; for the higher ranges, such as 100 and 1000 V, attenuation would be used. To make the DMM's "front end" useful, it must also act as an impedance-matching device. An ideal DMM has infinite input impedance and will not load the circuit being measured. As a practical matter, an op amp with a FET input stage will produce impedances of  $10^{10} \Omega$ . For the higher ranges where an attenuator is necessary, an input impedance of 10 M $\Omega$  is common. Figure 23.2 illustrates a typical dc input amplifier/attenuator.

Gains are selected by precision resistors in the feedback loop of the op amp. By changing the value of the resistors, the gain may be changed according to the following expression:  $(R_1 + R_2)/R_2$ . The precision resistors used are selected for excellent temperature tracking characteristics, and many are produced by using thin-film technology.

One of the sources of error in a DMM is the offset of the op amp. This directly affects accuracy and ability to measure small voltages. An automatic-zero circuit is commonly employed to overcome this problem, as illustrated in Fig. 22.3a and b. The dc offset is measured, stored, and then subtracted from the reading. The technique relies on the fact that



**Fig. 23.2** The dc input section to a DMM consists of an op amp with the gain controlled by precision resistors  $R_1$  and  $R_2$ . An attenuator is switched in for voltage ranges higher than the capability of the op amp.



**Fig. 23.3** The automatic-zero circuit cancels the effect of dc offsets by storing the offset on a capacitor during intervals when the input amplifier is not in use for a measurement. The stored offset is subtracted from the reading by placing the capacitor in series with the feedback resistors.

the input amplifier is used only during part of the measurement sequence. During auto-zero, the input is short-circuited to ground by switch S2. Switch S1 opens so that the input to the instrument from the user is not short-circuited to ground.

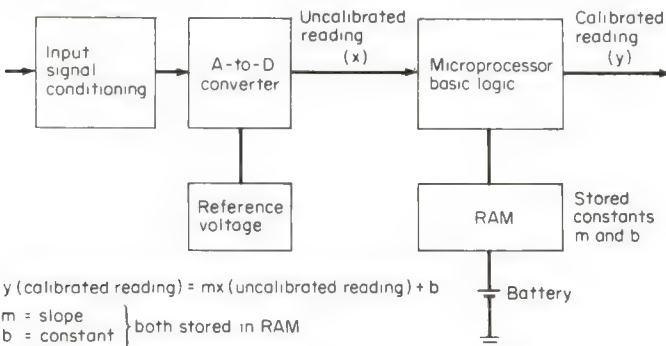
Any offset is stored on capacitor C relative to ground. When the amplifier is used for a measurement, switches S2, S3, and S5 are opened. The capacitor with the stored offset is introduced in series with the feedback loop but in opposite polarity. The capacitor acts as a battery, nulling the offset voltage.

The entire automatic-zero function can be handled digitally also. This has become quite practical through the use of a microprocessor. Many modern DMMs accomplish automatic-zeroing by short-circuiting the input between normal measurement cycles, measuring the value of the offset, and storing the value in random-access memory (RAM). By using the computational power of the microprocessor, the stored offset is subtracted from the subsequent measurement mathematically.

### 23.3 INTERNAL REFERENCE VOLTAGE

Many DMMs employ a zener reference diode as an internal voltage standard. Measurements are calibrated against this internal reference. A resistive divider allows the reference voltage from the zener diode to be calibrated periodically. The diode used is carefully selected for low temperature coefficient, and some instruments have the reference diode in a temperature-controlled chamber.

Some of the newer DMMs using microprocessors have eliminated all internal adjustments and employ electronic calibration, as shown in Fig. 23.4. During calibration the voltage standard is applied to the input, and the DMM is placed in a calibration routine. The calibration constants needed to produce a correct reading relative to the internal zener refer-



**Fig. 23.4** Electronic calibration makes use of a microprocessor controller to read and store the calibration constants necessary to produce correct readings. The reference voltage within the DMM need only be stable. This technique eliminates the need for internal adjustments.

ence diode are stored in a nonvolatile memory. During normal use, the calibration constants for that particular range and function are recalled from memory and applied mathematically to compute the correct reading. The nonvolatile memory typically consists of a CMOS RAM powered by a lithium battery with a life expectancy of up to 10 years.

### 23.4 AC CONVERSION

There are vast differences among ac converters used in DMMs. The ideal converter measures the rms value of the incoming signal and converts it to its equivalent dc voltage for digitization by the A/D converter. By definition, the rms value is the square root of the sum of the squares of the components of the ac signal, or the equivalent dc voltage required to produce the same heating effect as the ac signal.

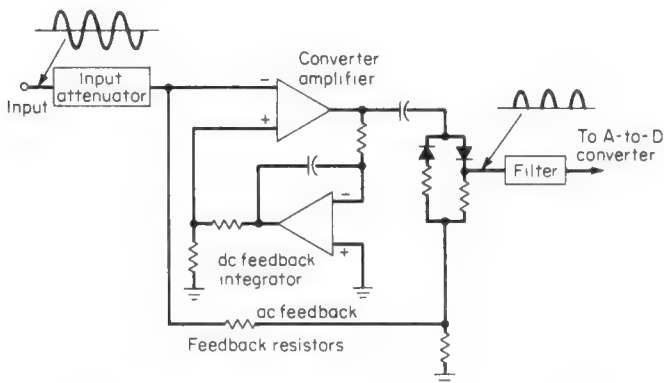
Some ac converters use a simple, inexpensive diode bridge, as shown in Fig. 23.5. The output from the bridge is filtered and scaled to produce a dc voltage proportional to the average value of the ac input. The scale factor selected applies to a pure sine wave (without harmonic content):

$$\frac{V_{rms}}{V_{avg}} = \frac{0.707}{0.637} = 1.11$$

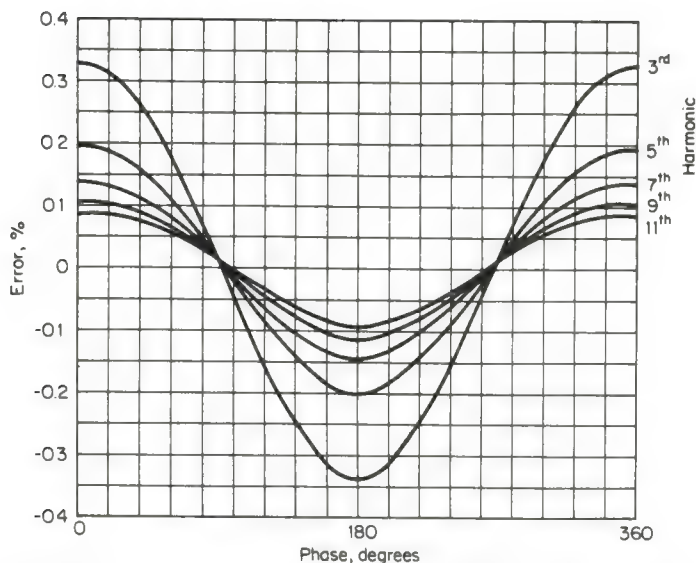
Should the odd harmonic distortion enter the signal, this scale factor changes (odd harmonics add an extra half-cycle of energy during the period of the fundamental). Even harmonics, however, cause little or no change in the scale factor to produce an rms reading. Since the scale factor is fixed, significant errors are produced by odd harmonic content. For an idea of just how severe the problem is, refer to Fig. 23.6 which shows percentage error versus phase for 1 percent odd harmonic content (relative to the fundamental) for average responding converters. The same type of information is illustrated in Fig. 23.7 for distortion as a function of amplitude.

Since even harmonics add as much as they subtract to the rectified waveform over the period of the fundamental, they contribute only small errors. For example, even harmonic content up to 10 percent produces an error below  $\pm 0.5$  percent.

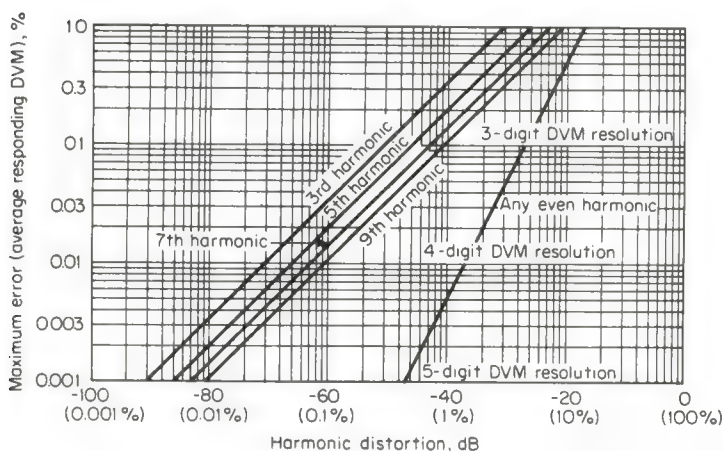
An ac converter that overcomes these limitations is a true rms converter. It responds to the dc heating value of the input signal independent of harmonic content up to the full bandwidth of the converter. One approach uses a combination thermocouple and heater and is illustrated in Fig. 23.8. The input signal is presented to the heater element, and the thermocouple measures the heating effect. The two components are in close thermal proximity. Since the dc output from the thermocouple is equivalent to the heating value of the incoming signal, by definition, this approach measures true rms.



**Fig. 23.5** Average responding ac converters use a diode bridge to rectify the signal. Filtering smooths the signal into a dc value equivalent to the rms value for a pure sine wave. Errors in this type of converter become significant when the input signal contains small amounts of odd harmonic distortion.

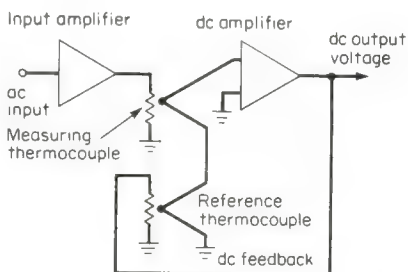


**Fig. 23.6** Percentage of error versus phase for 1 percent odd harmonic distortion is shown for an average responding ac converter.



**Fig. 23.7** Percentage of error as a function of odd harmonic distortion for a worst-case phase relationship with the fundamental.

**Fig. 23.8** True rms ac converters use a combination thermocouple and heater. The ac signal is presented to the heater element, and the thermocouple converts the heating effect to its dc equivalent voltage, which is digitized. A reference thermocouple is necessary to linearize the output and compensate for changes in ambient temperature.



Thermocouples are nonlinear as a function of input signal amplitude. They also respond to any change in ambient temperature. To make the thermocouple-heater approach useful, a second matched thermocouple operated in opposition is necessary to linearize the output and compensate for changes in ambient temperature. The sensitivity of the single thermocouple to small changes in the input signal can be improved by adding more junctions in series, but all in proximity to the same heater element. This is called a *thermopile*, and such devices are manufactured by using thin-film technology to reduce thermal inertia, thereby improving response time.

As a side note, thermocouples and thermopiles are square-law devices. Their output decreases at a rate much greater than proportional changes in the level of the input signal. For this reason, typically true rms converters are not specified for operation below one-tenth of full scale.

There is a third type of converter used in DMMs which has become quite popular. It is called a *quasi-rms*, or *computing rms*, converter. It simulates true rms action by using a series of op amps to square, take the average of the square, and then take the square root. In other words, it mathematically models true rms action. Since the synthesis is not perfect, such converters are limited to symmetric waveshapes. Square waves, triangular waves, and sine waves with up to 3 percent odd harmonic distortion can be measured effectively.

A figure of merit called the *crest factor* tells the user how much of a pulse can be measured by a given ac converter. It is defined as the peak value divided by the true rms value, or in terms of duty cycle, as follows:

$$\text{Crest factor} = \frac{V_{\text{peak}}}{V_{\text{rms}}} = \sqrt{\frac{1-D}{D}}$$

where  $D = \text{duty cycle} = T_1/(T_1 + T_2)$ .



Crest factors of 5:1 to 7:1 can be handled by quasi-rms converters while 10:1 crest factors are commonly specified for true rms converters.

**example 23.2** For third-harmonic distortion with a  $140^\circ$  phase shift with respect to the fundamental, what is the reading error on an average responding ac converter?

**solution** Using Fig. 23.6, we find the error is  $-2.5$  percent.

**example 23.3** If a true rms voltmeter could accurately measure a pulse with a 10 percent duty cycle, what would its crest factor be rated at?

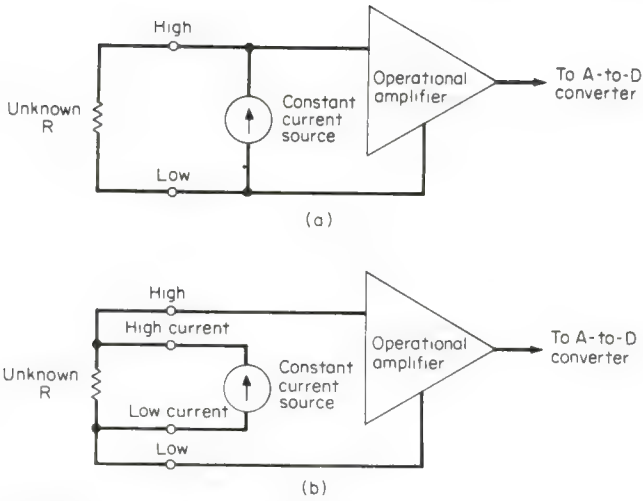
**solution** The square root of  $0.9/0.1$  is a crest factor of 3:1.

## 23.5 MEASURING RESISTANCE

The basic approach used to measure the value of an unknown resistor is to supply a constant dc current through the unknown and then, by using the DMM dc input, to measure the voltage drop across the unknown. The value of the resistor can be computed given the amount of current supplied and the voltage reading. The simplest approach, called a two-wire measurement, is shown in Fig. 23.9a.

When a DMM is used in a system where wire runs from the instrument are often long, the lead wire resistance introduces an error, especially for low-resistance readings. For example, a 6-ft wire can have  $0.060 \Omega$  of lead resistance. If a measurement is being made on a  $100\text{-}\Omega$  resistor with a five-digit DMM, the lead resistance will produce a 60-count error. A four-wire ohm converter solves this problem. As shown in Fig. 23.9b, sense wires which carry virtually no current to the DMM are used for the measurement. A separate pair of wires is used to supply current to the unknown resistor.



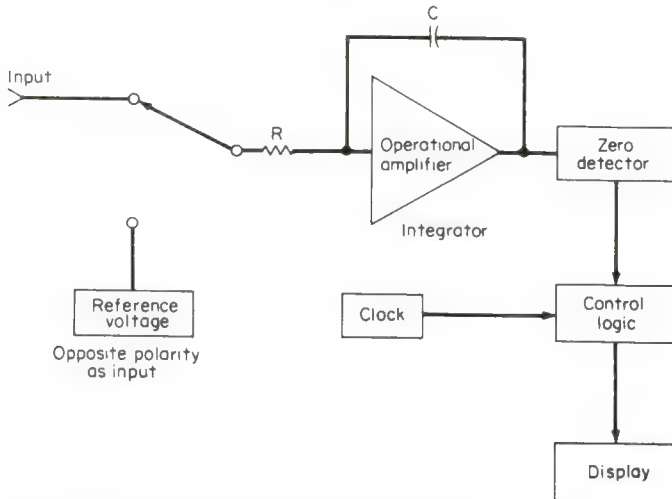


**Fig. 23.9** In systems applications, simple two-wire ohm converters (a) have yielded to four-wire converters (b) to eliminate errors caused by lead resistance.

### 23.6 A/D CONVERSION TECHNIQUES

The A/D converter in a DMM governs some of the instrument's basic characteristics, including accuracy, speed, and noise rejection. A/D converters may be divided into two major classes: integrating and nonintegrating. The integrating techniques mathematically reduce the effect of power-line-related noise. Nonintegrating techniques offer advantages in speed and broadband noise rejection when they are combined with filtering.

One of the most important and popular A/D conversion techniques is dual-slope integration. Owing to its simplicity, it is widely used; a block diagram is shown in Fig. 23.10.



**Fig. 23.10** The dual-slope A/D conversion technique uses an integrator as its key element. This integrating technique offers high noise rejection for line-related frequencies.



The input voltage is connected to an integrator via switch  $S$  for the duration of the integration period. Integration periods are selected to be related to the line frequency period; i.e., for 60-Hz line frequency,  $\frac{1}{60}$  and  $\frac{1}{120}$ -s integration periods are common. During this time interval,  $C$  is charged at a rate governed by  $R$ . At the end of the integration period, the charge on  $C$  is proportional to the input voltage. The integrator action of a capacitor in the feedback loop of an op amp causes the voltage to charge in a linear fashion. The charging rate is governed by the current through  $R$ .

In the second phase of dual-slope integration,  $S$  is connected to a reference voltage of the opposite polarity as the input voltage. Then  $C$  is discharged linearly by the reference voltage until zero crossover is reached. At the beginning of the discharge cycle, a clock is gated on and counts are accumulated until zero crossover. The total count at the end of the measurement period is proportional to the input voltage and is scaled and displayed as the reading.

If the input voltage is doubled, as illustrated in Fig. 23.11, then  $C$  charges twice as fast to twice the value. When the reference voltage is applied, it takes twice as long to discharge  $C$ , and the accumulated counts are, therefore, doubled.

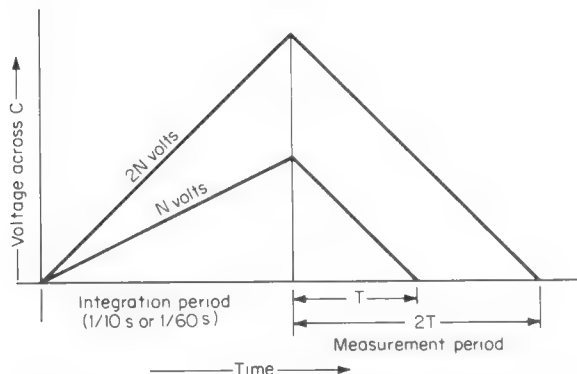
Considerable time can be saved during the measurement cycle if the reading is developed during discharge. An obvious solution is to increase the discharge slope to speed up the reading rate. As the slope increases, however, being able to accurately detect zero crossover becomes more difficult. A shallow slope produces more accurate readings. A modification to dual slope can be made by using a steep slope initially; then just prior to zero crossover, the slope is changed to a shallow one.

As microprocessors dropped in price and became more practical for instrumentation, a variation on the dual-slope one called a multislope DMM came into being. The multislope employs a steep rundown initially, as illustrated in Fig. 23.12. A finite amount of overshoot is allowed beyond the zero crossover. A slope exactly a decade less in steepness is applied next to the remaining voltage until a second zero crossover occurs, also with a certain amount of overshoot. At this point, a third slope yet another decade less in steepness is applied. Finally, a fourth slope one-thousandth as steep is applied for the final zero crossover.

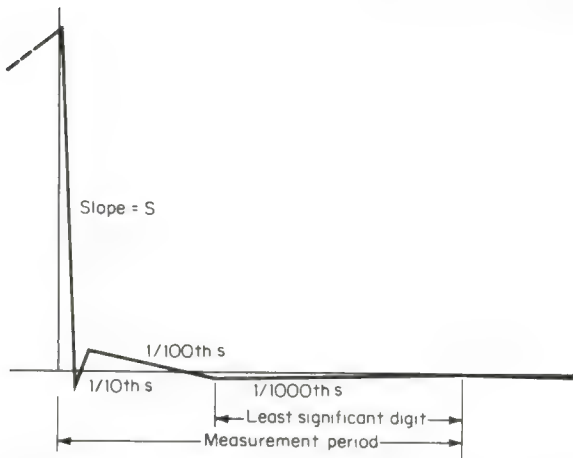
At the beginning of each slope, a clock is started and then stopped at the zero crossover. The accumulated clock pulses for each slope represent the four least significant digits in the reading (10 000 counts). The digits are developed as the discharge cycle takes place, improving the instrument's reading speed.

Note that the steepest slope (with the least definition in its zero crossover) is the least critical to the reading. The shallowest slope with the greatest definition in zero crossover is used to determine the most sensitive digit.

Multislope DMMs can offer a tradeoff in resolution versus reading speed by reducing the integration period and terminating the reading during the rundown period for the desired



**Fig. 23.11** The voltage on the integrator capacitor  $C$  is directly proportional to the magnitude of the input voltage. This example shows two different input voltages, one twice the magnitude of the other. During the measurement period, the number of counts accumulated at zero crossover is exactly double.



**Fig. 23.12** The multislope integrating technique is an expansion of the dual-slope technique. The least significant digits in the reading are developed during the rundown period. Slopes exactly a decade less in steepness are successively applied to the voltage on the integrating capacitor. A certain amount of overshoot is allowed each time until the final zero crossover occurs.

number of digits. Fewer digits require less time to develop. Table 23.1 shows an example of the specifications for a specific commercially available unit. The tradeoff is simple: the greatest resolution and sensitivity are achieved at the greatest sacrifice in reading speed.

During the integration period, the most significant digits can be developed. With the simple dual-slope technique, the input voltage is allowed to ramp during the integration period in a linear fashion. For very small input voltages, the charge remaining on the integrator capacitor is proportionately less and, therefore, susceptible to noise. A sawtooth approach can be used during the integration period to fully utilize the dynamic range of the integrator. The sawtooth is created by allowing the input voltage to ramp up; then by switching in the reference voltage (of opposite polarity), the voltage level is forced to ramp down. At a predetermined level, the input voltage is applied again, the voltage ramps up, and so on until the end of the integration period. The number of transitions necessary during the integration period required to keep the voltage within range of the integrator is used to develop the two most significant digits. When they are combined with the digits developed during the discharge period, an accurate six-digit reading can be developed.

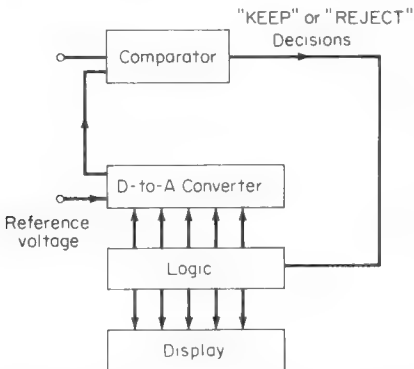
Successive approximation represents one of the most important nonintegrating A/D techniques. A succession of level changes is generated by a D/A converter and compared with the input voltage, as shown in Fig. 23.13. If the ultimate reading is to have binary-coded decimal (BCD) weighting, then the D/A converter produces a sequence of voltages which have the same weighting (8:4:2:1). Other weighting schemes may be employed also.

For each digit, four "tries," or approximations, are made until the number of desired digits is developed. The comparator is used to decide whether the try is to be retained as part of the final reading or rejected. Accuracy is limited by the comparator, and the tech-

**TABLE 23.1 Reading Speed of Multislope DMM**

	Reading rate per second		
	Three digits	Four digits	Five digits
Automatic-zero off	71	33	4.4
Automatic-zero on	53	20	2.3

nique is quite vulnerable to noise. It is simple and extremely fast, making it popular for systems applications in which reading speed is important.

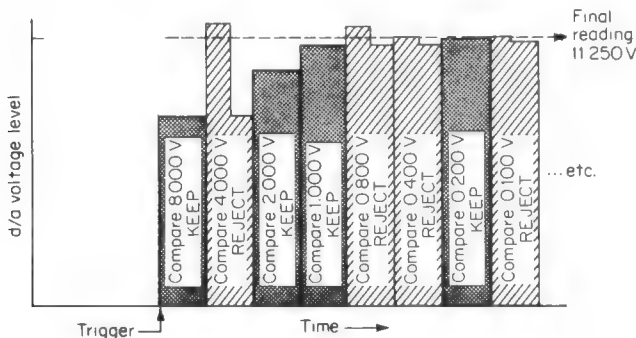


**Fig. 23.13** Successive approximation is a technique used in DMMs where reading speed is important. A succession of level changes are generated by a D/A converter and compared with the input voltage to derive the final reading.

As an example of how successive approximation works, assume that the full-scale input to the A/D converter is 15 V and that an 11.250-V signal is applied. The first digit is established through the generation of 8, 4, 2, and 1 V. Refer to Fig. 23.14. The comparator keeps the 8 V, rejects the combination of 8 V and 4 V, keeps the combination of 8 and 2 V, and keeps the combination of 8, 2, and 1 V. Then the magnitude of the voltage levels is dropped by a decade to produce 0.8, 0.4, 0.2, and 0.1 V. The 11 V from the first approximation is combined sequentially with the second approximation. The combination of 11 and 0.8 V is rejected, the combination of 11 and 0.4 V is also rejected, but the combination of 11 and 0.2 V is kept. The combination of 11.2 and 0.1 V is rejected. After application of a sequence a decade lower (0.08, 0.04, 0.02, and 0.01 V), the final reading is derived as the sum of 8, 2, 1, 0.2, 0.04, and 0.01 V, or 11.250 V.

**example 23.4** What is the greatest advantage of an integrating DMM?

**solution** Rejection of power-line-related noise.



**Fig. 23.14** This is an example of how successive approximation works for digitizing an input voltage equal to 11.250 V. The comparator either keeps or rejects the voltage generated for each "try," or approximation. Through a series of successive tries the final reading is developed.

## Chapter 24

# Oscilloscope Measurements

**Chet Heyberger**

*Portable Instruments Div.  
Tektronix, Inc.  
Beaverton, Oregon*

### 24.1 INTRODUCTION\*

Almost anything that can be measured, can be measured with the two-dimensional graph drawn by an oscilloscope. In most applications the oscilloscope shows a graph of voltage on the vertical axis versus time on the horizontal axis. This general-purpose display presents far more information than is available from other test and measurement instruments such as frequency counters or multimeters. For example, with an oscilloscope you can determine how much of a signal is direct current, how much is alternating current, how much is noise, whether the noise is changing with time, and what the frequency of the signal is. Using an oscilloscope lets you see everything at once rather than through many separate tests.

The usefulness of an oscilloscope is broad because most electric signals can be connected easily to the oscilloscope with either probes or cables. And to measure nonelectrical phenomena, transducers are available. Transducers, such as speakers and microphones, change one kind of energy into another. Other typical transducers can transform mechanical stress, pressure, light, or heat to electric signals.

The oscilloscope is a true multipurpose instrument. The front panel controls allow you to pick from a wide range of vertical sensitivities, time references, display modes, and triggering possibilities.

In other words, given the proper transducer and a modern oscilloscope, your test and measurement capabilities are almost limitless. To tap that measurement potential, you need to know how to use an oscilloscope. This chapter provides basic information on oscilloscope measurement techniques. It is divided into two main parts describing first the controls of an oscilloscope and then some common measurement techniques. In concluding sections we describe the effects of the scope's performance on your measurements; storage oscilloscopes, specialized instruments that offer measurement solutions not found anywhere else; and how to select an oscilloscope.

Wherever it is necessary for the sake of clarity to refer to an actual instrument instead of oscilloscopes in general, either the Tektronix 2213 or the 2215 portable oscilloscope is used as an example.

\*This chapter has been adapted from the Tektronix publication *The XYZs of Using a Scope*, copyright 1981, with the permission of Tektronix, Inc., Beaverton, Oregon.

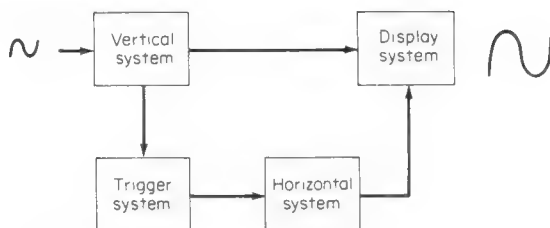


Fig. 24.1 Block diagram of a basic oscilloscope.

## 24.2 OSCILLOSCOPE CONTROLS

Making oscilloscope measurements will be easier if you understand the basics of how an oscilloscope works and which controls enable you to perform the operations you want. With that goal in mind, this section describes the controls of a triggered-sweep, dual-channel, general-purpose oscilloscope.

Think of the oscilloscope in terms of the functional blocks illustrated in Fig. 24.1; these include vertical, trigger, horizontal, and display systems. The vertical system controls the vertical axis of the display. Any time the electron beam that draws the graph moves up or down, it does so under control of the vertical system. The horizontal system controls the left-to-right movement of the beam. The trigger system determines the precise moment when the oscilloscope draws a display by “triggering” the beginning of the horizontal sweep across the screen. The display system contains the cathode-ray tube (CRT) on which the graph is drawn.

**DISPLAY SYSTEM** The oscilloscope draws a graph by moving an electron beam across a phosphor coating on the inside of the CRT. The result is a glow for a short time afterward, tracing the path of the beam. A grid of lines etched or silkscreened on the inside of the faceplate serves as the reference for your measurements; this is the graticule shown in Fig. 24.2.

The graticule is a grid of lines typically etched or silkscreened on the inside of the CRT faceplate. Putting the graticule inside—on the same plane as the trace drawn by the electron beam, and not on the outside of the glass—eliminates measurement inaccuracies called *parallax* errors. Parallax errors occur when the trace and the graticule are on different planes and the observer is shifted slightly from the direct line of sight. Although different-sized CRTs may be used, graticules are usually laid out in an  $8 \times 10$  pattern. Each of the 8 vertical and 10 horizontal lines blocks off major divisions (or just divisions) of the screen. The labeling on the oscilloscope controls always refers to major divisions. The tick marks on the center graticule lines represent minor divisions, or subdivisions. Since risetime measurements are very common, most oscilloscope graticules include risetime measurement mark-

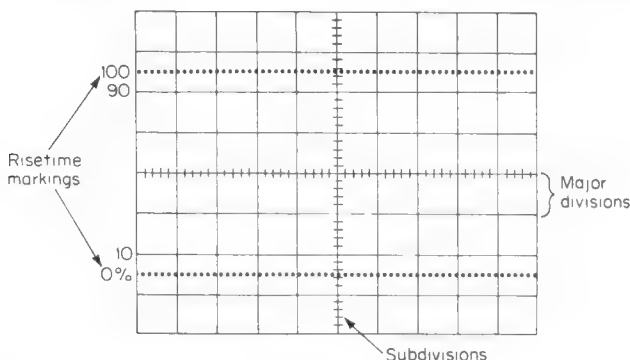


Fig. 24.2 Graticule of an oscilloscope.



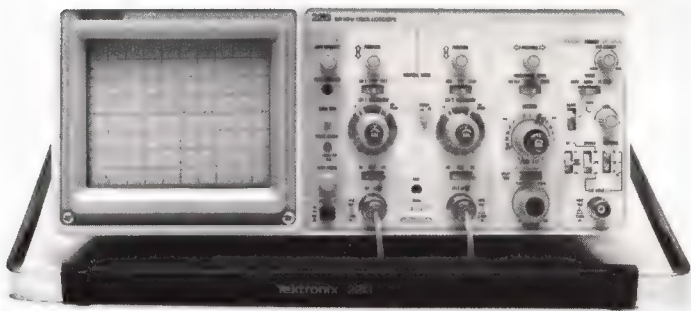


Fig. 24.3 Front panel of a dual-channel 60-MHz oscilloscope.

ings: dashed lines for 0 and 100 percent points and labeled graticule lines for 10 and 90 percent points.

Common controls for display systems include intensity and focus; less often, you will find beam finder and trace rotation controls. The functions of these controls are described below, and their positions on the Tektronix 2213 portable oscilloscope are illustrated in Fig. 24.3. A functional diagram of the display section is shown in Fig. 24.4.

The display system of an oscilloscope consists of the CRT and its controls. To draw a graph of the measurements, the vertical system of the scope supplies the Y, or vertical, coordinates, and the horizontal system supplies the X, or horizontal, coordinates. There is also a Z axis in an oscilloscope; it determines whether the electron beam is turned on and how bright it is.

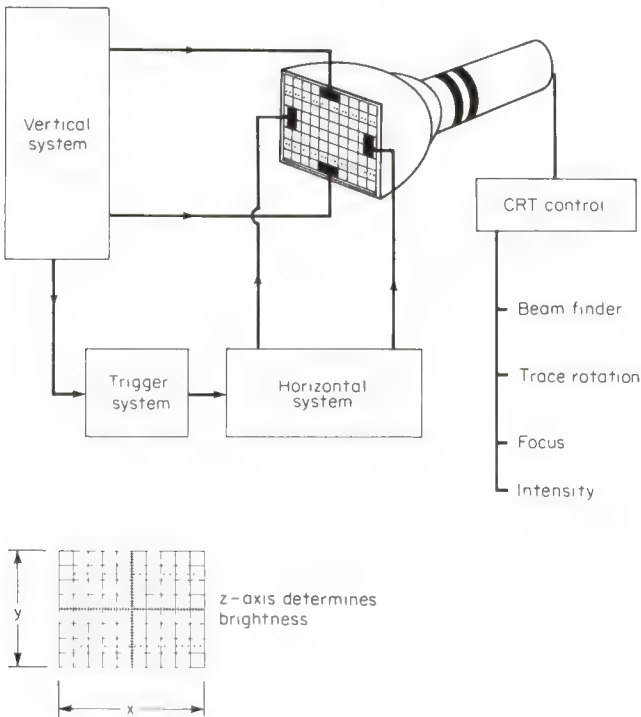


Fig. 24.4 Display system and controls of an oscilloscope.



**Beam finder** The beam finder is a convenience control that allows you to locate the electron beam any time it is off screen. When you push the BEAM FIND button, you reduce the vertical and horizontal deflection voltages (more about deflection voltages later) and override the intensity control, so that the beam always appears within the  $8 \times 10$  cm screen. When you see which quadrant of the screen the beam appears in, you know which direction to turn the horizontal and vertical position controls to reposition the trace for normal operations.

**Intensity** An intensity control adjusts the brightness of the trace. It is necessary because you use an oscilloscope in different ambient light conditions and with many kinds of signals. For instance, with square waves you might want to change the trace brightness because the slower horizontal components always appear brighter than the faster vertical components.

Intensity controls are also useful because the intensity of a trace is a function of two things: how bright the beam is, and how long it is on screen. As you select different sweep speeds (a sweep is one movement of the electron beam across the oscilloscope screen) with the SEC/DIV switch, the beam on and off times change, and the beam has more or less time to excite the CRT's phosphor.

**Focus** The oscilloscope's electron beam is focused on the CRT faceplate by an electrical grid within the tube. The focus control adjusts that grid for optimum trace focus.

**Trace rotation** Another display adjustment on the front panel of some oscilloscopes is trace rotation. This control allows you to electrically align the horizontal deflection, or baseline, of the trace with the fixed graticule lines. To avoid accidental misalignments when the oscilloscope is in use, this control is often recessed and must be adjusted with a small screwdriver.

Normally trace rotation can be set once in most oscilloscope applications, but the earth's magnetic field affects the trace alignment. And when an oscilloscope is used in many different positions, as a service oscilloscope will be, it is very handy to have a front panel trace rotation adjustment.

**VERTICAL SYSTEM** The vertical system of the oscilloscope supplies the display system with the  $y$ , or vertical axis of information for the graph on the CRT screen. To do this, the vertical system uses the input signals to develop deflection voltages. Then the display system uses these deflection voltages to control or deflect the electron beam.

The vertical system also gives you a choice of how you connect the input signals (called *coupling* and described later). And the vertical system provides internal signals for the trigger circuit (described later). Figure 24.5 illustrates the vertical system schematically.

The vertical system of a dual-channel oscilloscope consists of two identical channels (although only one is shown in Fig. 24.5). Each channel has circuits to couple an input signal to that channel, attenuate (reduce) the input signal when necessary, preamplify it, delay it, and finally amplify the signal for use by the display system. The delay line lets you see the beginning of a waveform even when the oscilloscope is triggering on it.

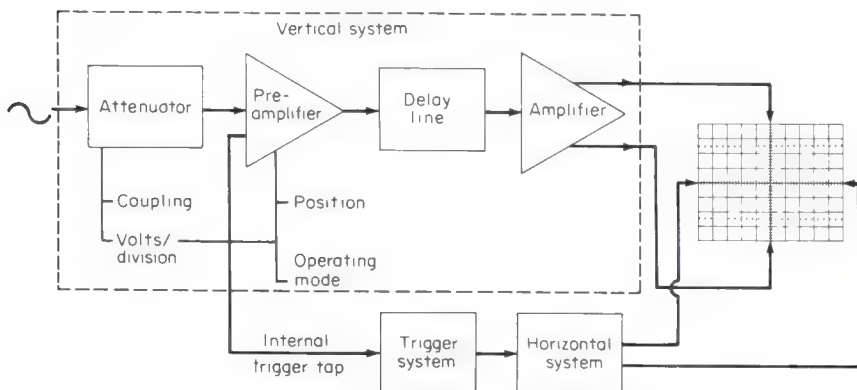


Fig. 24.5 Vertical system of a dual-channel (only one is shown here) oscilloscope.

**Vertical position** The oscilloscope's position controls let you place the trace exactly where you want it on the screen. The two vertical position controls (there's one for each channel in a dual-trace oscilloscope) change the vertical placement of the traces from each vertical channel; the horizontal position control changes the horizontal position of both channels at once.

**Input coupling** The input coupling switch for each vertical channel lets you control how the input signal is coupled to the vertical channel. The direct-current (dc) input coupling lets you see an entire input signal. Alternating-current (ac) coupling blocks the constant or dc signal components and permits only the ac components of the input signal to reach the channel.

The middle position of the coupling switches is marked GND, for ground. Choosing this position disconnects the input signal from the vertical system and shows a trace representing the oscilloscope's chassis ground if the scope is triggered. The position of the trace on the screen in this mode is the ground reference level. Switching from AC or DC to GND and back is a handy way to measure signal voltage levels with respect to chassis ground. Note that using the GND position does not ground the circuit you are measuring.

**VOLTS/DIV switch** The rotary VOLTS/DIV switch controls the sensitivity of each vertical channel. Having different sensitivities extends the range of the oscilloscope's applications. With a VOLTS/DIV switch, a multipurpose oscilloscope is capable of accurately displaying input signal levels from millivolts to many volts.

Using the VOLTS/DIV switch to change sensitivity also changes the scale factor, the value of each major vertical division of the graticule. Each setting of the VOLTS/DIV control is marked with a number that represents the scale factor for that channel. For example, with a setting of 10 V each of the eight vertical major divisions represents 10 V, and the entire screen can show 80 V from bottom to top. With a VOLTS/DIV setting of 2 mV, the eight vertical divisions can display 16 mV.

The probe you use influences the scale factor. Note that there are two marked areas under the skirts of the VOLTS/DIV switches on most Tektronix scopes. The right-hand area shows the scale factor when you use a standard 10X probe. The left-hand area shows the factor for a 1X probe.

**Variable volts per division** The red CAL (for calibration) control in the center of the VOLTS/DIV switch provides a continuously variable change in the scale factor to a maximum greater than 2.5 times the VOLTS/DIV setting. The control is labeled CAL to remind you that only one position of this control is calibrated; that switch position usually has a detent.

A variable sensitivity control is useful when you want to make quick amplitude comparisons on a series of signals. You could, for example, take a known signal of almost any amplitude and use the CAL control to make sure the waveform fits exactly on major division graticule lines. Then as you use the same vertical channel to look at other signals, you can quickly see whether the later signals have the same amplitude.

**Channel 2 inversion** To make differential measurements (described later), you have to invert the polarity of one of your input channels. The INVERT control on the vertical amplifier for channel 2 provides this capability. When you push it in, the signal on channel 2 is inverted. When the switch is out, both channels have the same polarity.

**Vertical operating modes** Oscilloscopes are more useful if they have more than one vertical display mode. With dual-channel oscilloscopes, the choices are usually channel 1 alone, channel 2 alone, both channels in either the alternate or chopped mode, and both channels algebraically summed.

Dual-channel oscilloscopes are more useful than single-channel instruments because they allow you to compare a known good signal on one channel with an unknown signal on the other. Two channels also allow you to compare two signals that are either related to or dependent on each other. Of course, you can still use a dual-channel oscilloscope one channel at a time.

Switches in the vertical system area of the front panel allow you to choose the vertical operating mode.

To make the oscilloscope display only channel 1, use the CH1 vertical-mode switch setting. To display only channel 2, use the CH2 setting. To display two separate signals, use either the ALT (alternate) or CHOP (chopped) mode.

Both CHOP and ALT modes are provided so that you can look at two signals at any sweep speed. The alternate mode draws first one trace and then the other, completing one before

starting the other. This works well at faster sweep speeds where your eyes cannot see the alternation. To see two signals at slower sweeps, use the chopped mode.

When you want to see two input signals combined into one waveform on the screen, use the ADD mode. This gives you an algebraically combined signal: either channels 1 and 2 added, CH 1 + CH 2; or channel 1 minus channel 2 when channel 2 is inverted, +CH 1 + (-CH 2).

**HORIZONTAL SYSTEM** To draw a graph, the oscilloscope needs horizontal as well as vertical data. The horizontal system supplies the second dimension by providing the deflection voltages to move the electron beam horizontally. The horizontal system contains a sweep generator that produces a sawtooth waveform, or ramp (see Fig. 24.6), that is used to control the oscilloscope's sweep rates.

The sweep generator makes the unique functions of the modern oscilloscope possible. The circuit that made the rate of rise in the ramp linear (a refinement pioneered by Tektronix) was one of the most important advances in oscillography. It meant that the horizontal beam movement could be calibrated directly in units of time. That advance makes it possible for you to measure time between events much more accurately on the oscilloscope screen.

Because it is calibrated in units of time, the sweep generator is often called the *time base*. It lets you pick the time units, observing the signal for either very short times (measured in nanoseconds or microseconds) or relatively long times of several seconds.

The sawtooth waveform is a voltage ramp produced by the sweep generator. The rising portion of the waveform is called the *ramp*, the falling edge is the *retrace*, and the time between ramps is the *holdoff* time. The sweep of the electron beam across the screen of an oscilloscope is controlled by the ramp, and the return of the beam to the left side of the screen takes place during the retrace.

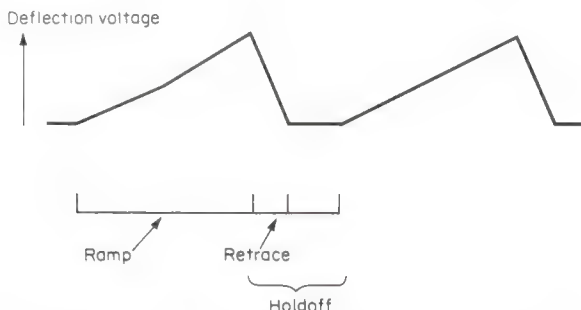
The two functional blocks of the horizontal system, the horizontal amplifier and sweep generator, are diagramed in Fig. 24.7. The sweep generator produces a sawtooth waveform that is processed by the amplifier and applied to the horizontal deflection plates of the CRT. The horizontal system also provides the *z* axis of the oscilloscope; the *z* axis determines whether the electron beam is turned on and how bright it is.

**Horizontal position** Like the vertical position controls, you use the horizontal position control to change the location of the waveforms on the screen.

**Seconds per division** The SEC/DIV switch lets you select the rate at which the beam sweeps across the screen; changing the SEC/DIV settings allows you to look at longer or shorter time intervals of the input signal. Like the vertical system VOLTS/DIV switch, the SEC/DIV markings refer to the screen's scale factors. If the SEC/DIV setting is 1 ms, each horizontal major division represents 1 ms, and the total screen shows you 10 ms.

On dual-time-base oscilloscopes like the Tektronix 2215, there are two SEC/DIV controls. The A sweep offers all the possible settings while the SEC/DIV switch for the delayed B sweep usually lacks one or two of the slower settings. (On the 2215, there are B sweep settings for 0.05  $\mu$ s to 50 ms per division.)

Oscilloscopes also have an XY setting on the SEC/DIV switch for making the XY measurements described later.



**Fig. 24.6** Sawtooth waveform (or ramp) that controls the horizontal sweep of an oscilloscope.

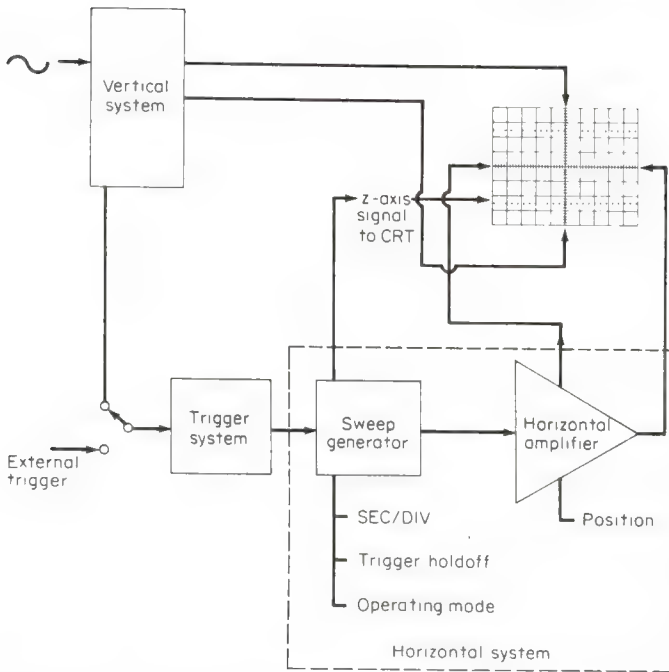


Fig. 24.7 Horizontal system of a dual-channel oscilloscope.

**Variable seconds per division** Besides providing calibrated speeds, most oscilloscopes allow you to change any sweep speed by turning the red CAL (calibrated) control in the center of the SEC/DIV switch counterclockwise. This control slows the sweep speed by at least 2.5:1, making the slowest sweep you have  $0.5 \times 2.5 = 1.25$  s per division. Remember that the detent (usually in the extreme clockwise direction) is the calibrated position.

**Horizontal magnification** Most oscilloscopes offer some means of horizontally magnifying the waveforms on the screen. The effect of magnification is to multiply the sweep speed by the amount of magnification. On 2200 series oscilloscopes, there is a 10X horizontal magnification that you engage by pulling out the red CAL switch. The 10X horizontal magnification gives you a sweep speed 10 times faster than the SEC/DIV switch setting; for example,  $0.05 \mu\text{s}$  per division magnified is 5 ns per division.

The 10X magnification is useful when you want to look at signals and see details that occur very closely together in time. The magnification control of an oscilloscope always gives you the fastest possible sweep speeds.

**Horizontal operating mode** Most general-purpose oscilloscopes do not have a horizontal operating mode control; they have a single time base and only one operating mode. But when an oscilloscope is capable of delayed-sweep measurements, there are horizontal operating mode options.

The front panels of such oscilloscopes are marked A, or NORM, for the normal undelayed sweep and B or DELAYED, for the delayed B sweep. Combinations of the two are possible also.

For more detail, see the description of delayed-sweep measurements under "Measurement Techniques."

**TRIGGER SYSTEM** So far you have found that the display system draws the waveforms on the screen, the vertical system supplies the vertical information for the drawing, and the horizontal system provides the time axis. At this point the only thing that is missing is the "when"—when should the other circuits of the oscilloscope start drawing the signal and when should they not?



The "when" is the trigger, and it is important for a number of reasons. First, getting time-related information is one reason you use an oscilloscope. Second, and equally important, each drawing starts with the same "when."

Each sweep will start at the right time if you make the right trigger system control settings. You tell the trigger circuit which trigger signal to select with the source switches. With an external signal, you connect the trigger signal to the trigger system circuit with the external coupling controls. Next you set the trigger circuit to recognize a particular voltage level on the trigger signal with the slope and level controls. Then every time that level occurs, the sweep generator is turned on. The process is diagrammed in Fig. 24.8.

Triggering gives you a stable display because the same trigger point starts the sweep each time. SLOPE and LEVEL controls define the trigger points on the trigger signal. When you look at a waveform on the screen, you see all those sweeps overlaid onto what appears to be one display.

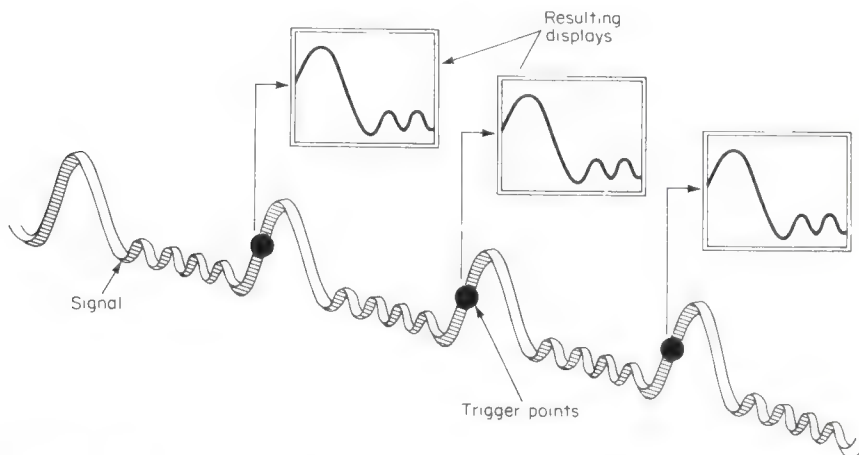
Besides the trigger controls already mentioned, you have controls that determine how the trigger system operates (trigger operating mode) and how long the oscilloscope waits between triggers (holdoff). On some dual-time-base scopes, there is a separate trigger for the B sweep.

The trigger circuit and its controls are shown schematically in Fig. 24.9. The trigger source control describes whether the trigger signal is internal or external to the oscilloscope. The coupling controls determine the connection of an external trigger to the trigger circuit. The level and slope controls determine where the trigger point will be on the trigger signal. And the trigger operating-mode control determines how the trigger circuit operates.

**Trigger level and slope** These controls define the trigger point. The slope control determines whether the trigger point is found on the rising or falling edge of a signal. It specifies either a positive (also called the rising, or positive-going) edge or a negative (falling, or negative-going) edge. The level control determines at what voltage level on that edge the trigger point occurs. See Fig. 24.10.

**Variable trigger holdoff** Not every trigger event can be accepted as a trigger. The trigger system will not recognize a trigger during the sweep or the retrace or for a short time afterward, called the holdoff period. The retrace is the time it takes the electron beam to return to the left side of the screen. The holdoff period provides additional time beyond the retrace to ensure that the oscilloscope is ready to start another sweep.

Sometimes the normal holdoff period is not long enough to ensure a stable display; this possibility exists when the trigger signal is a complex waveform with many possible trigger points on it. Although the waveform is repetitive, a simple trigger might yield a series of patterns on the screen instead of the same pattern each time. Digital pulse trains are a good example; each pulse is very much like any other, so there are many possible trigger points, not all of which result in the same display.



**Fig. 24.8** How specific trigger points produce the resultant displays on the CRT.

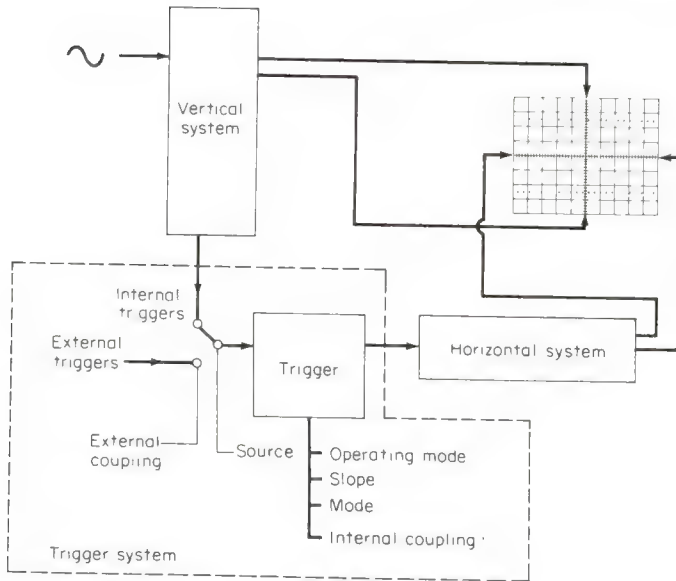


Fig. 24.9 Trigger system for a dual-channel oscilloscope.

What you need now is some way to control when a trigger point is accepted. The variable trigger holdoff control provides the capability. (The control is actually part of the horizontal system—because it adjusts the holdoff time of the sweep generator—but its function interacts with the trigger controls.) Figure 24.11 diagrams a situation in which the variable holdoff is useful. The variable holdoff control lets you make the oscilloscope ignore some potential trigger points. In the figure, all the possible trigger points in the input signal would result in an unstable display. Changing the holdoff time to make sure that the trigger point appears on the same pulse in each repetition of the input signal is the only way to ensure a stable waveform.

**Trigger sources** Trigger sources are grouped into two categories that depend on whether the trigger signal is provided internally or externally. The source makes no difference in how the trigger circuit operates, but internal triggering usually means the oscilloscope is triggering on the same signal that it is displaying. That has the obvious advantage of letting you see where you are triggering.

Triggering on the displayed signal is not always what you need, so external triggering is also available. It often gives you more control over the display. To use an external trigger, you set the SOURCE switch on its EXT (external) position and connect the triggering signal to the input connector marked EXT INPUT on the front panel. Occasions when external triggering is useful often occur in digital design and repair; you might want to look at a long

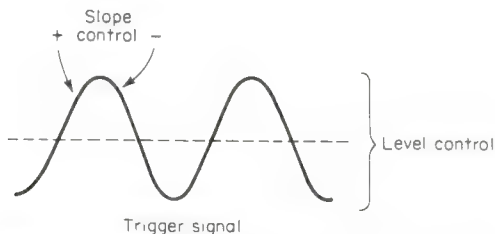


Fig. 24.10 Effect of the slope and level controls on the trigger signal.



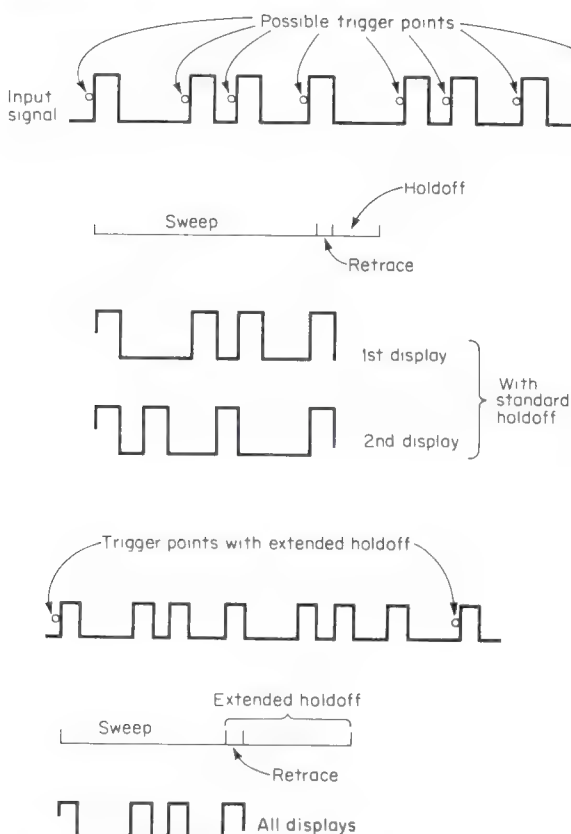


Fig. 24.11 Effect of the variable trigger holdoff control.

train of very similar pulses while triggering with an external clock or with a signal from another part of the circuit.

The line position on the source switch gives you another triggering possibility: the power line. Line triggering is useful any time you are looking at circuits that depend on the power-line frequency. Examples include devices such as light dimmers and power supplies.

**Trigger operating modes** Trigger circuits of oscilloscopes like the Tektronix 2213 and 2215 can operate in four modes: normal, automatic, television, and vertical.

One of the most useful is the normal trigger mode (usually marked NORM on the front panel) because it can handle a wider range of trigger signals than any other triggering mode. The normal mode does not permit a trace to be drawn on the screen if there is no trigger. The normal mode usually gives you the widest range of triggering signals.

In the automatic, or "bright baseline," mode (labeled AUTO on front panels), a trigger starts a sweep, the sweep ends, and the holdoff period expires; at that point a timer begins to run. If another trigger is not found before the timer runs out, a trigger is generated anyway, causing the bright baseline to appear even when there is no waveform on the channel. The automatic mode lets you trigger on signals with changing voltage amplitudes or wave-shapes without adjusting the LEVEL control.

Another useful operating mode is television triggering. Most oscilloscopes with this mode let you trigger on TV fields at sweeps of 100  $\mu$ s per division and slower and TV lines at 50  $\mu$ s per division or faster. Others can trigger on either fields or lines at any sweep speed.

You will probably use the normal and automatic modes the most, the AUTO mode because it is essentially totally automatic and the normal mode because it is the most ver-

**TABLE 24.1** Types of Trigger Couplings

Coupling	Description
Direct current	Direct current couples all elements of the triggering signal (both ac and dc) to the trigger circuit.
Direct current with attenuation	If you want dc coupling and the external trigger is too large for the trigger system, move the trigger coupling switch to its $DC \div 10$ setting.
Alternating current	This coupling blocks dc components of the trigger signal and couples only the ac components.
High-/low-frequency rejection	These couplings prevent the passage of either high or low frequencies in the triggering signal. You use them to prevent triggering on parts of the waveform that would result in unstable displays.

satellite. For example, it is possible to have a low-frequency signal with a repetition rate that is mismatched to the runout of an automatic-mode timer. When that happens, the signal will not be steady in the automatic mode. Moreover, automatic modes are often signal-seeking and cannot trigger on very low-frequency trigger signals. The normal mode, however, gives you a steady signal at any repetition rate.

A less common trigger circuit operating mode is the alternate, or vertical, mode (vertical mode on the Tektronix 2213 and 2215). This operating mode has unique advantages; the oscilloscope triggers alternately on the two vertical channels. Thus you can look at two completely unrelated signals. Most oscilloscopes trigger on only one channel or the other when the two signals are not synchronous.

**Trigger coupling** Just as you may select either ac or dc coupling when you connect an input signal to your oscilloscope's vertical system, you can select the kind of coupling you need when you connect a signal to the trigger system. Some frequently offered choices are shown in Table 24.1.

## 24.3 OSCILLOSCOPE MEASUREMENTS

In previous sections we described the oscilloscope controls used to select the exact oscilloscope functions you need to make the measurements you want. Now we discuss the procedures and techniques you will need when actually making measurements.

**SAFETY** Before you make any oscilloscope measurement, remember that you must be careful when you work with electric equipment. Always observe all safety precautions described in the operator's or service manual for the equipment you are using. Some general rules about servicing electric equipment are worth repeating here. Do not service electric devices alone. Know the symbols for dangerous circuits, and observe the safety instructions for your equipment. Do not operate an electric device in an explosive atmosphere. Always ground the oscilloscope to the circuit, and ground both the oscilloscope and the circuit under test. (Remember that if you lose the ground, all accessible conductive parts, including knobs that appear to be insulated, can give you a shock.) To avoid personal injury, do not touch exposed connections and components in the circuit under test when the power is on. And remember to consult the service manual for the equipment you are working on.

There are also rules about the oscilloscope itself. To avoid a shock, plug the power cord of the oscilloscope into a properly wired receptacle before connecting your probes; use only the power cord for your oscilloscope, and do not use one that is not in good condition (if it is cracked or broken, is missing a ground pin, etc.). Use the right fuse to avoid fire hazards. Do not remove covers and panels on your oscilloscope.

**GETTING SET UP** Accurate oscilloscope measurements require that you make sure your system is properly set up each time you begin to use your oscilloscope. A very common and frustrating source of inaccurate oscilloscope measurements is the result of forgetting to check the controls to make sure they are where you think they are. Check the oscilloscope's controls before you begin making a measurement. Use the functional sections of the instrument as the outline for your checklist.

1. Check the settings on all the vertical system controls: variable controls (CH 1 and CH 2, CAL, VOLTS/DIV) should be in their calibrated detent positions; make sure CH 2

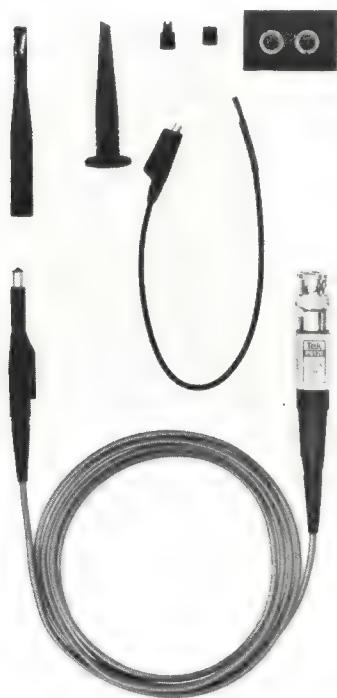
is not inverted (unless you want it to be). Check the vertical-mode switches to make sure the signal from the proper channel(s) will be displayed. Check the two vertical system VOLTS/DIV switches to see whether their settings are right (and do not forget to use the VOLTS/DIV readout that matches the probe, either 1X or 10X. Check the input coupling levers, too.

2. Check the horizontal system control settings: magnification is off (push in the red CAL switch in the middle of the SEC/DIV switch), and variable SEC/DIV is in its calibrated detent position.

3. Check your trigger system controls to make sure the oscilloscope will pick the right slope on the trigger signal, that the right coupling is selected, and that the correct operating mode will be used. Also make sure that the trigger variable holdoff control is at its minimum position.

4. Adjust the display section controls (focus, intensity, trace rotation, beam finder), if necessary.

**PROBES** Connecting measurement test points to the inputs of your oscilloscope is best done with a probe like the one illustrated in Fig. 24.12. Probes are the usual connection



**Fig. 24.12** An oscilloscope probe and its accessories.

test. Tektronix probes consist of a patented resistive cable and a grounded shield. Probes like the Tektronix P6120 are usually supplied when you purchase an oscilloscope. The P6120 probe is a high-impedance, minimum-loading 10X passive probe. The accessories pictured with it include (from left to right) a grabber tip for integrated circuits (ICs) and small-diameter leads, a retractable hook tip, an IC tester tip cover, an insulating ground cover, marker bands, and the ground lead in the center. All are worthwhile additions to your test and measurement system.

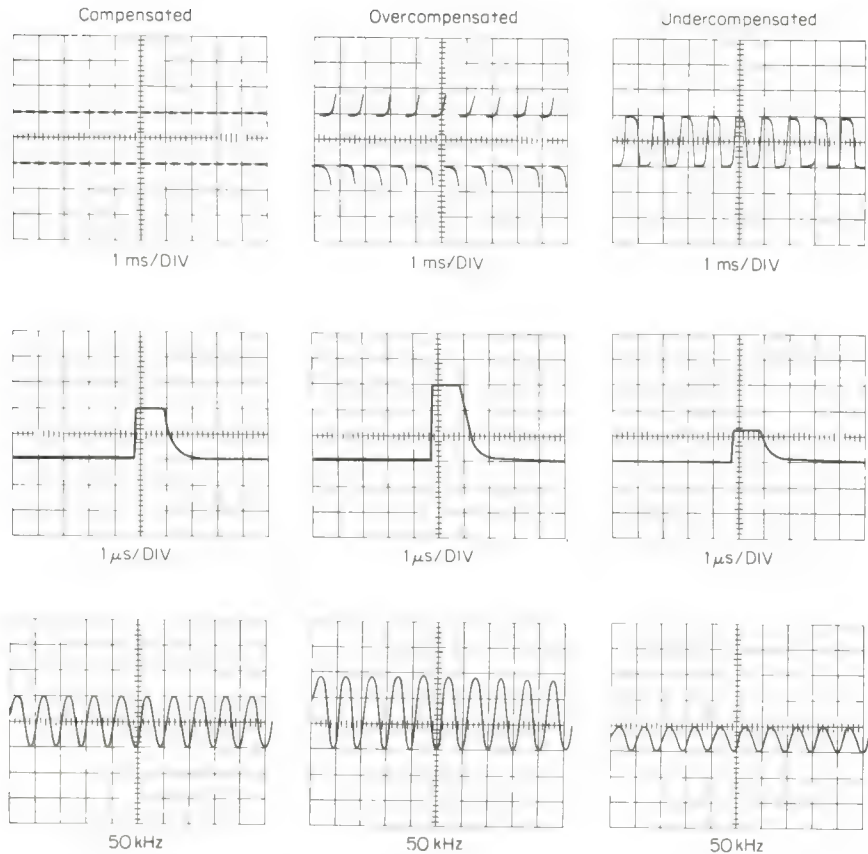
**Probe compensation** Most measurements you make with an oscilloscope require an attenuator probe, which is any probe that reduces signal amplitude. The most common are 10X passive probes which reduce the amplitude of the signal by a factor of 10:1.

Before you make any measurement with an attenuator probe, you should make sure it is compensated. Compensation matches that particular probe to the inputs of the oscilloscope you are using. To make compensation easy, oscilloscopes have a compensation signal output jack on their front panels, usually marked CAL, CAL OUTPUT, or PROBE ADJUST.

Figure 24.13 illustrates what can happen to the waveforms you measure when the probe is not properly compensated. Improperly compensated probes can distort the waveforms you see on the screen. In the figure, a square wave, a pulse, and a sine wave are shown as they will appear with proper and improper compensations.

Note that you should compensate your probe as it will be used when you make the measurement: compensate it with the accessory tip you will be using. Do not compensate the probe in one vertical channel and then use it on another.

**Probe handling** Before you probe a circuit, make sure you have the right probe tips and adaptors for the circuits you will be working on. Then make sure that the ground in the circuit under test is the same as the scope ground—do not just assume it is. The oscil-



**Fig. 24.13** Effects of properly and improperly compensated oscilloscope probes.

oscope ground will always be earth ground as long as you are using the proper power cord and plug. Check the circuit ground by touching the probe tip to the point you think is ground before you make a "hard" ground by attaching the ground strap of your probe.

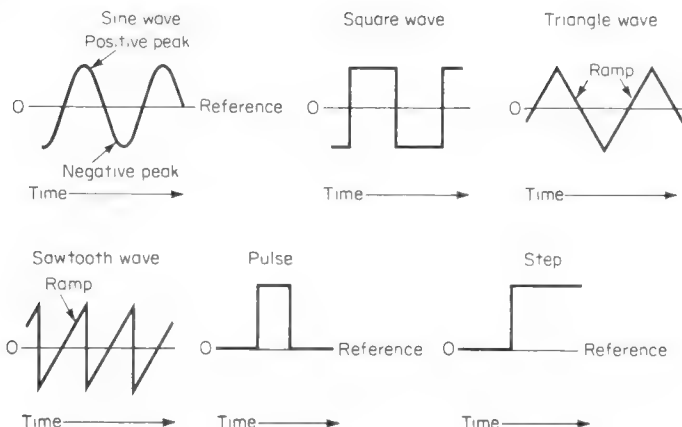
If you are going to be probing a lot of different points in the same circuit to measure low frequencies, you can ground that circuit to the oscilloscope once instead of each time you move the probe. Connect the circuit ground to the jack marked GND on the front panel (if this jack was provided by the oscilloscope manufacturer).

**BASIC WAVEFORMS** What you will actually see on the screen of an oscilloscope is called a *waveform*, a graphic representation of an electric signal. The changes in the waveform with time form the waveshape, the most readily identifiable characteristic of a waveform. Figure 24.14 illustrates some common waveshapes.

Basic waveshapes include sine waves and various nonsinusoidal waves such as triangular waves, rectangular waves like the square wave, and sawtooth waves. A square wave has equal amounts of time for its two states. Triangular and sawtooth waves are usually the result of circuits designed to control voltage with respect to time, such as the sweep of an oscilloscope and some television circuits. In these waveforms, one or both transitions from state to state are made with a steady variation at a constant rate, a ramp. (Changes from one state to another on all waveforms except sine waves are called transitions.) The last two drawings represent aperiodic, single-shot waveforms. The first is a pulse; all pulses are marked by a rise, a finite duration, and a decay. The second one is a step, which is actually a single transition.

Waveshapes tell you a great deal about the signal. As a matter of fact, before oscilloscopes

## 24-14 Oscilloscope Measurements



**Fig. 24.14** Some common waveforms that may be seen on an oscilloscope.

had a linear time base for making time measurements, they were still acceptable design and service instruments because so much information is contained in just the shape of a signal.

In examining waveshapes, remember that a change in the amplitude, or vertical dimension, of a signal represents a change in signal voltage. Any time a flat horizontal line is displayed, there was no change for that length of time. Straight diagonal lines indicate a linear change, equal rise or fall of voltage for equal amounts of time. Sharp angles on a waveform indicate a sudden change.

But waveshapes alone are not the whole story. To completely describe an electric signal, you will want to find the parameters of that particular waveform. Depending on the signal, these parameters might be amplitude, period, frequency, width, risetime, or phase. You can review these signal parameters and the necessary measurements in the following paragraphs.

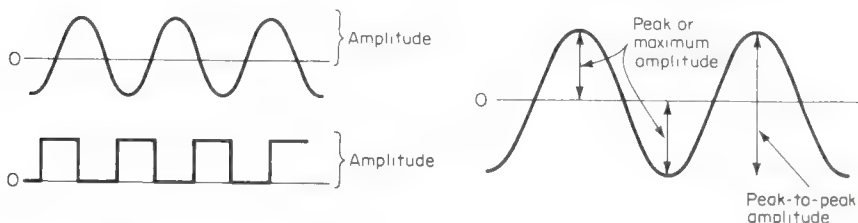
**MEASUREMENT TECHNIQUES** Rather than attempt to describe how to make all the possible measurements, in this section we describe common measurement techniques you will use in almost every application.

**Direct and derived measurements** The two most basic measurements you can make are the two direct oscilloscope measurements: amplitude and time. Almost every other measurement is based on one or both of these direct oscilloscope measurements.

Since the oscilloscope is a voltage-measuring device, voltage is shown as amplitude on the screen. Of course, voltage, current, resistance, and power are related:

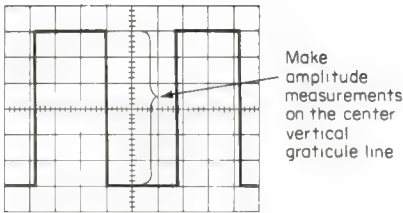
$$\begin{aligned}\text{Current} &= \frac{\text{voltage}}{\text{resistance}} \\ \text{Resistance} &= \frac{\text{voltage}}{\text{current}} \\ \text{Power} &= \text{current} \times \text{voltage}\end{aligned}$$

Figure 24.15 illustrates the amplitude parameter. Amplitude is a characteristic of all waveforms. It is the amount of displacement from equilibrium at a particular point in time.

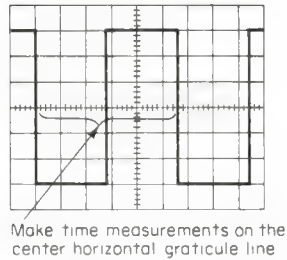


**Fig. 24.15** Amplitude parameter of two waveforms.





**Fig. 24.16** Method of making an amplitude measurement by using the graticule.



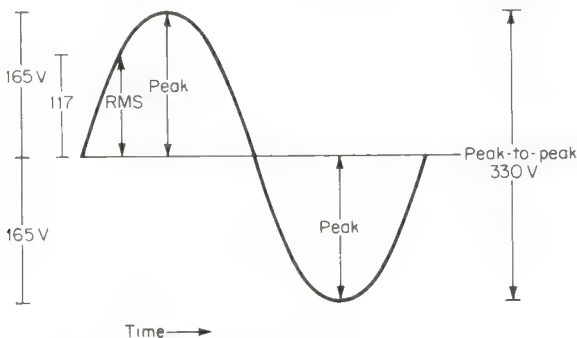
**Fig. 24.17** How to make a time measurement by using the graticule.

Note that without a modifier, the word means the maximum change from a reference without regard to the direction of the change. In the first two drawings (sine wave and square wave), the amplitude is the same even though the sine wave is larger from peak to peak. In the third drawing, an ac waveform is shown with peak (or maximum) amplitude and peak-to-peak amplitude parameters annotated. In oscilloscope measurements, amplitude usually means peak-to-peak amplitude.

Amplitude measurements are best made with a signal that covers most of the screen vertically such as in Fig. 24.16. Time measurements are also more accurate when the signal covers a large area of the screen; see Fig. 24.17. Use the oscilloscope's vertical position control to place the waveform on a major division of the graticule. Then use the horizontal position control to make the lowest horizontal component of the waveform intersect the center vertical graticule line. With a display like that in the figure, you can count major and minor divisions down the center vertical graticule line and multiply by the VOLTS/DIV setting to make the measurement. For example, 6.0 divisions times 0.1 V equals 0.6 V.

Time measurements are also more accurate when the signal covers a large area of the screen (see Fig. 24.17). Use the horizontal position control to place a vertical portion of the waveform on a major division of the graticule. Then count major and minor divisions across the center horizontal graticule line to make your time measurement. Multiply by the SEC/DIV setting; for example, 5.3 divisions times 0.2 ms equals 1.06 ms.

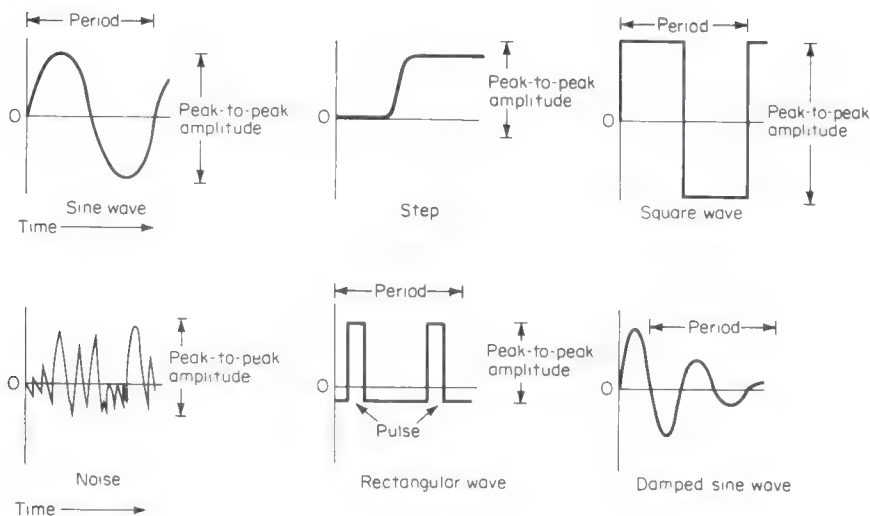
Derived measurements are any measurements that are calculated from direct measurements; see Fig. 24.18. For example, ac measurements require an amplitude measurement first. The easiest place to start is with a peak-to-peak amplitude measurement of the voltage—in this case, 330 V because peak-to-peak measurements ignore plus and minus signs. The peak voltage is one-half that value (when there is no dc offset) and is also called a maximum value; it is 165 V in this case. The average value is the total area under the voltage curves divided by the period in radians; for a sine wave, the average value is zero because the positive and negative values are equal. The root mean square (rms) voltage for this sine wave, which represents the line voltage in the United States, is equal to the maximum value divided by  $\sqrt{2}$ :  $165/1.414 = 117$  V.



**Fig. 24.18** A derived measurement.



## 24-16 Oscilloscope Measurements



**Fig. 24.19** Period measurements.

**example 24.1** Use the derived measurements described in Fig. 24.18, and compute the peak-to-peak, peak, and rms voltage of the waveform.

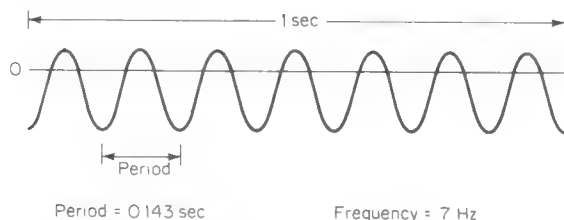
**solution** Since peak-to-peak (p-p) voltage is measured without regard to positive or negative, the waveform is 1 V p-p. The peak voltage is  $-1$  V because the waveform is negative-going. The rms voltage is 0.707 V.

**Period measurements** One of the most frequently used parameters of a signal is the period; see Fig. 24.19. The technique used to measure period is a direct time measurement. Period is the time required for one cycle of a signal if the signal repeats itself. Period is a parameter whether the signal is symmetrically shaped like the sine and square waves in the figure or whether it has a more complex and asymmetric shape like the rectangular wave and damped sine wave. Period is always expressed in units of time. One-time signals such as the step or signals without a time relation such as noise have no period.

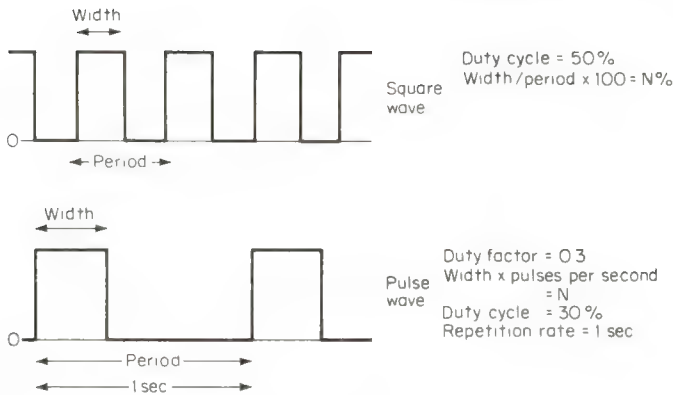
**example 24.2** If the oscilloscope showing the waveform in Fig. 24.19 has a SEC, DIV switch setting of 0.5 ms, what is the period of the waveform?

**solution** The waveform takes five major divisions on the screen to repeat itself, so its period is  $5 \times 0.5 \text{ ms} = 2.5 \text{ ms}$ .

**Frequency measurements** Frequency is an example of a derived measurement. One technique requires first making a direct period measurement. While period is the time required to complete one cycle of a periodic waveform, frequency is the number of cycles that take place in 1 s. The measurement unit is 1 Hz (1 cycle per second), and it is the reciprocal of the period. So a period of 0.00114 s (or 1.14 ms) means a frequency of 877 Hz. Figure 24.20 contains an illustration.



**Fig. 24.20** A frequency measurement.



**Fig. 24.21** Rectangular-wave measurements.

**example 24.3** Calculate the frequency of the waveform pictured in Example 24.2.

**solution** Period and frequency are reciprocal, so a period of 2.5 ms means a frequency of 400 Hz ( $1/0.0025$ ).

Another method of making a frequency measurement is described under “XY Measurements.”

**Rectangular-wave measurements** Rectangular waves—square waves are the most common example—have many parameters, as illustrated in Fig. 24.21. The parameters described are all derived from direct time measurements.

Duty cycle, duty factor, and repetition rate are parameters of all rectangular waves. They are particularly important in digital circuitry. Duty cycle is the ratio of pulse width to signal period, expressed as a percentage. For square waves, the ratio is always 50 percent, as you can see; for the pulse wave in the second drawing, it is 30 percent. Duty factor is the same thing as duty cycle except it is expressed as a decimal, not a percentage. A repetition rate describes how often a pulse train occurs and is used instead of frequency to describe waveforms like that in the second drawing.

**example 24.4** Use the measurements described and the waveform below to calculate duty cycle, duty factor, and repetition rate. The instrument’s SEC/DIV setting is 1 ms.

**solution** Duty cycle and duty factor depend on the ratio of pulse width to period. Without referring to the instrument settings, you can simply count major divisions:

$$\begin{aligned} \text{Duty cycle:} & \quad \% \times 100 = 75\% \\ \text{Duty factor:} & \quad \% = 0.75 \end{aligned}$$

To find the repetition rate, you need the SEC/DIV setting. Note that the waveform repeats itself in four major divisions:  $4 \times 1 \text{ ms per division} = 4 \text{ ms}$ .

**Pulse measurements** Pulse measurements are important when you work with digital equipment and data communications devices. Some of the signal parameters of a pulse are shown in Fig. 24.22, an illustration of an ideal pulse. The most important parameters of a real pulse are shown in Fig. 24.23.

Pulse parameters can be important in a number of different applications. (See Fig. 24.22.) Digital circuitry, x-ray equipment, and data communications are examples. Pulse specifications include transition times measured on the leading edge of a positive-going transition; this is the risetime. The falltime is the transition time on a negative-going trailing edge. Pulse width is measured at the 50 percent points and amplitude from 0 to 100 percent. Any displacement from 0 V for the base of the pulse is the baseline offset.

The measurement of real pulses includes a few more parameters than those for an ideal pulse. In Fig. 24.23, several are shown. Pres shooting is a usually small change of amplitude preceding and in the opposite direction to the pulse rise. Overshooting and rounding are changes that occur after the initial transition. Ringing is a series of amplitude changes—usually a damped sinusoid—that follow overshooting. All are expressed as percentages of

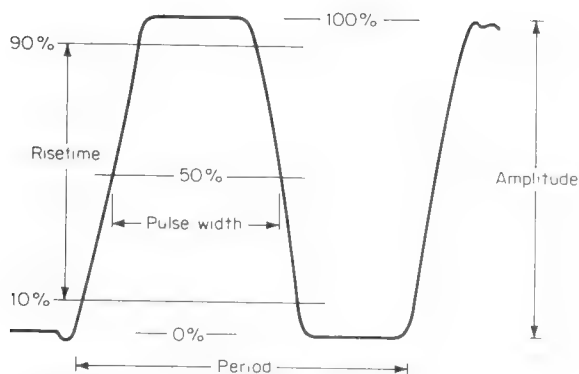


Fig. 24.22 Some signal parameters of an ideal pulse.

amplitude. Droop is a decrease in the maximum amplitude with time. And nonlinearity is any variation from a straight line drawn through the 10 and 90 percent points of a transition.

**example 24.5** What are the risetime and pulse width of the waveform pictured if the oscilloscope's SEC/DIV switch setting is 0.2 ms?

**solution** The risetime is measured from the 10 to the 90 percent point of the rising edge of the waveform. It takes two major divisions for the waveform to cross those points, so the risetime is  $2 \times 0.2 \text{ ms} = 0.4 \text{ ms}$ . A pulse width is always measured at the 50 percent point of a pulse; here the waveform takes six major and four minor divisions:  $6.8 \times 0.2 \text{ ms} = 1.36 \text{ ms}$ . (Note that each minor division is 0.2 of a major division.)

**Phase measurements** One of the characteristics of an ac waveform is phase. Phase is the amount of time that has passed since the cycle began, measured in degrees, where one complete cycle is  $360^\circ$ . This parameter is illustrated by Fig. 24.24.

There is also a phase relationship between two or more waveforms, as illustrated by Fig. 24.24. There are two ways to measure the phase shift between two waveforms. One way is to put one waveform on each channel of a dual-channel oscilloscope and view them directly in the chopped or alternate vertical mode. Trigger on either channel; then adjust the trigger LEVEL control for a stable display and measure the period of the waveforms. Next increase the sweep speed so that you have a display something like the second drawing in Fig. 24.24. Measure the horizontal distance between the same points on the two waveforms. The phase shift is the difference in time divided by the period and multiplied by  $360^\circ$ :

$$\theta = \frac{\Delta T}{P} \times 360^\circ$$

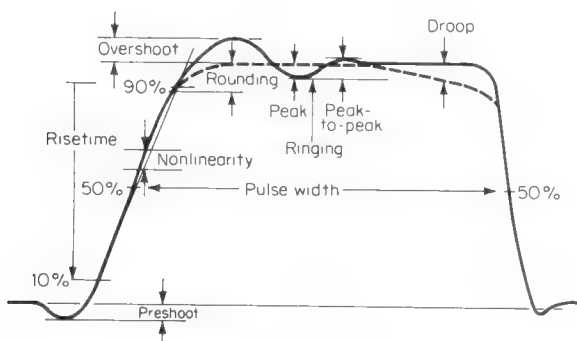
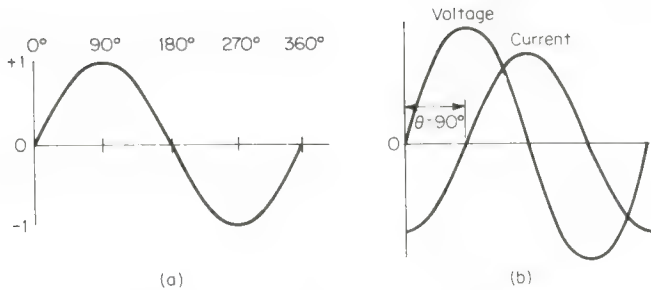


Fig. 24.23 Signal parameters present in a real pulse.



**Fig. 24.24** Phase measurement as seen on a dual-channel oscilloscope.

Phase is best explained with a sine wave. Remember that this waveform is based on the sine of all the angles from  $0^\circ$  through  $360^\circ$ . The result is a plot that changes from 0 at  $0^\circ$ , 1 at  $90^\circ$ , 0 again at  $180^\circ$ , -1 at  $270^\circ$ , and to 0 again at  $360^\circ$ . Consequently, it is useful to refer to the phase angle (or simply phase, when there is no ambiguity) of a sine wave when you want to describe how much of the period has elapsed.

Another use of phase is found when you want to describe a relationship between two signals. Picture two clocks with their second hands sweeping the dial every 60 s. If the second hands touch the 12 at the same time, the clocks are said to be *in phase*; if the hands do not touch, the clocks are *out of phase*. How far out of phase they are can be expressed as degrees of phase shift. To illustrate, the waveform labeled CURRENT in Fig. 24.24 is said to be  $90^\circ$  out of phase with the voltage waveform. Another way of reporting the same information is to say that the current waveform has a  $90^\circ$  phase angle with respect to the voltage waveform or that the current waveform lags the voltage waveform by  $90^\circ$ . Note that there is always a reference to another waveform; in this case, between the voltage and current waveforms of an inductor.

**example 24.6** If two sine waves have the same period of 12 ms but one starts 0.5 ms later, what is the phase shift of the second sine wave?

**solution** Phase shift is the difference in time divided by the period and multiplied by  $360^\circ$ :

$$\frac{0.5}{12} \times 360^\circ = 15^\circ$$

**XY measurements** Displaying two periodic ac waveforms and measuring when one starts with respect to another are possible with any dual-trace oscilloscope, but that is not the only way to make a phase measurement. Look at the front panel, and you will see that the vertical channel input connectors are labeled X and Y. The last position on the SEC/DIV switch is XY; and when you use it, the oscilloscope's time base is bypassed. The channel 1 input signal becomes the horizontal axis of the oscilloscope's display, and the signal on channel 2 becomes the vertical axis. In the XY mode, you can input one sinusoidal waveform on each channel, and your screen will display what is called a *Lissajous pattern*. The shape of the pattern will indicate the phase difference between the two signals. Some examples of Lissajous patterns are shown in Fig. 24.25.

Note that general-purpose oscilloscope Lissajous pattern phase measurements are usually limited by the frequency response of the horizontal amplifier (typically designed with far less bandwidth than vertical channels). Specialized XY oscilloscopes or monitors will have almost identical vertical and horizontal systems.

Finding the phase shift of two sinusoidal signals with a Lissajous pattern is one example of an XY measurement. The XY capability can be used for other measurements as well. The Lissajous patterns also can be used to determine the frequency of an unknown signal when you have a known signal on the other channel. This is a very accurate frequency measurement as long as your known signal is accurate and both signals are sine waves. The patterns you can see are illustrated in Fig. 24.25, where the effects of both frequency and phase differences are shown.

Frequency measurements with Lissajous patterns require a known sine wave on one channel. If there is no phase shift, the ratio between the known and unknown signals will correspond to the ratio of horizontal and vertical lobes of the pattern. When the frequencies

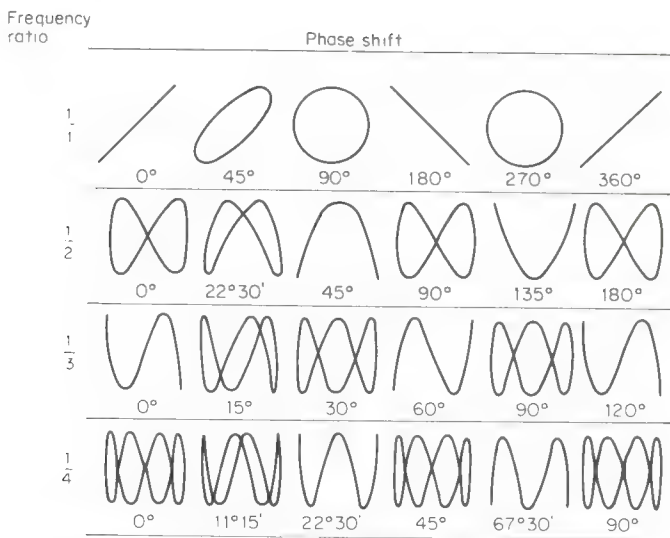


Fig. 24.25 Examples of Lissajous patterns.

are the same, only the shifts in phase will affect the pattern. In Fig. 24.25, both phase and frequency differences are shown.

There are many other applications for *XY* measurements in TV servicing, component checking, engine analysis, and two-way radio servicing, for example. In fact, any time you have physical phenomena that are interdependent and not time-dependent, *XY* measurements are the answer. With the proper transducer, you can use an oscilloscope's *XY* capability to make measurements of aerodynamic lift and drag, motor speed and torque, or pressure and volume of liquids and gases.

**Differential measurements** The ADD vertical mode and the channel 2 inversion switch of your oscilloscope allow you to make differential measurements. Often differential measurements let you eliminate undesirable components from a signal that you are trying to measure. If you have a signal that is very similar to the unnecessary noise, the setup is simple. Put the signal with the spurious information on channel 1. Connect the signal that is like the unwanted components to channel 2. Set both input coupling switches to DC (use AC if the dc components of the signal are too large), and select the alternate vertical mode.

Set your VOLTS/DIV switches so that the two signals are about equal in amplitude. Then you can change the vertical operating-mode switch to ADD and press the inversion switch, so that the common-mode signals have opposite polarities.

If you use the channel 2 VOLTS/DIV switch and CAL control for maximum cancellation of the common-mode signal, the signal that remains on screen will contain only the desired

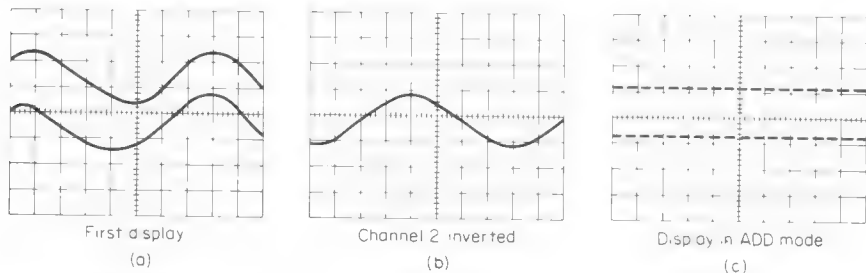


Fig. 24.26 Differential measurement.

part of the channel 1 input signal. The two common-mode signals have canceled, leaving only the difference between the two.

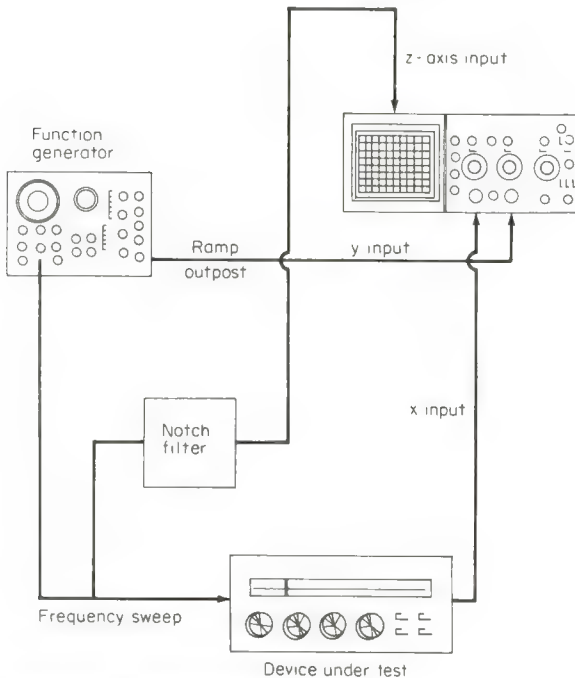
Differential measurements allow you to remove unwanted information from a signal any time you have another signal that closely resembles the unwanted components. (See Fig. 24.26.) For example, the first drawing shows a square wave contaminated by a sine wave. Once the common-mode component (the sine wave) is inputted to channel 2 and that channel is inverted, the signals can be added with the ADD vertical mode. The result is shown in the second drawing.

**Z axis measurements** Remember that the CRT in the oscilloscope has three axes of information: X is the horizontal component of the display, Y is the vertical, and Z is the brightness of the electron beam. When an oscilloscope has an external Z axis input connector on the back of the instrument, you can change the brightness (modulate the intensity) of the signal on the screen with an external signal. The Z axis input usually will accept a signal of up to 30 V where positive voltages decrease the brightness and negative voltages increase it; usually 5 V will cause a noticeable change.

The Z axis input is an advantage to users that have their instruments set up for a long series of tests. One example is the testing of high-fidelity equipment illustrated by Fig. 24.27.

Using the Z axis can provide additional information on the oscilloscope screen. In the setup shown in Fig. 24.27, a function generator sweeps through the frequencies of interest during the product testing, 20 to 20 000 Hz in this case. Then an adjustable notch filter is used to generate a marker, at 15 kHz, for instance, and this marker signal is applied to the Z axis input to brighten the trace. This setup allows the tester to evaluate the product's performance at a glance.

**Delayed-sweep measurements** Delayed sweep is a technique that adds a precise amount of time between the trigger point and the beginning of a horizontal sweep. Delayed sweep is often used as a convenient way to make a measurement, but sometimes it is the only way to make a measurement. Suppose, for example, that the part of the waveform you want to measure is so far from the only available trigger point that it will not show on the



**Fig. 24.27** Use of the Z axis of a CRT.



screen. The problem can be solved with delayed sweep. Trigger where you have to, and add delay out to where you want the sweep to start. The delayed-sweep feature that makes this easy is the intensified sweep mode; it lets you use the delayed sweep as a positionable magnifier.

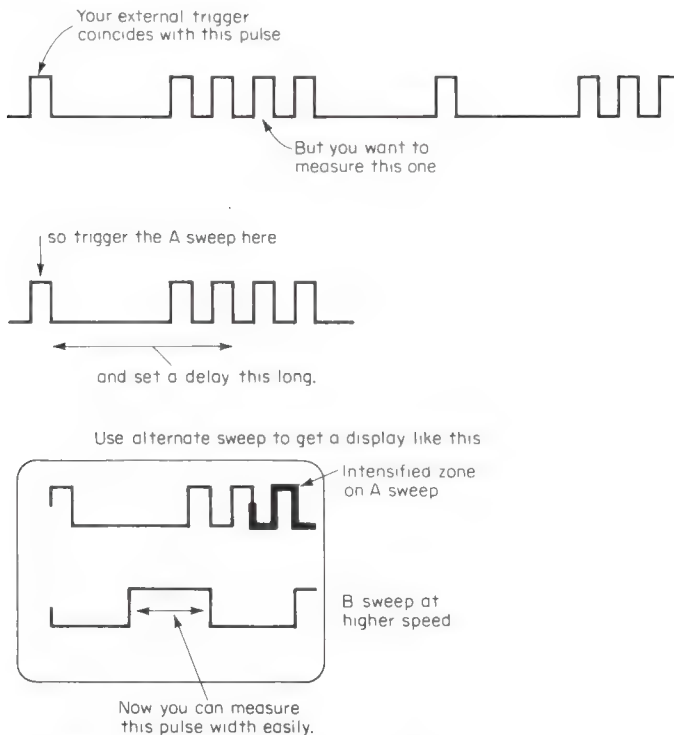
You trigger normally and then switch to the oscilloscope's intensified sweep mode. Now the signal on the screen will show a brighter zone after the delay time. Run the delay time (and the intensified zone) out to the part of the signal that interests you. Then switch to the delayed mode and increase the sweep speed to magnify the selected waveform portion so that you can examine it in detail. Figure 24.28 illustrates the process. The drawings above show a common measurement problem solved with the use of delayed sweep. The trigger event in this case is too far from the area of interest on the signal. Adding a delay between the trigger event and the start of the oscilloscope sweep is the answer. The intensified sweep feature of the oscilloscope makes setting the correct delay time very simple.

**Dual-Time-Base Oscilloscopes.** Delayed sweep is normally found on dual-time-base oscilloscopes like the Tektronix 2215; these instruments have two totally separate horizontal sweep generators. In dual-time-base instruments, one sweep is triggered in the normal fashion, and the start of the second sweep is delayed. To distinguish between these two sweeps, the delaying sweep is called the *A sweep* and the delayed sweep is called the *B sweep*. The time between the start of the A sweep and the start of the B sweep is called the *delay time*.

Dual-time-base oscilloscopes offer you all the measurement capabilities of single-time-base instruments plus the following:

- Convenient comparisons of signals at two different sweep speeds
- Jitter free triggering of delayed sweeps
- Timing measurement accuracy of 1.5 percent

Most of this increase in measurement performance is available because you can separately control the two sweep speeds and use them to examine two time slices—one large and one small—of the same signal.



**Fig. 24.28** A measurement problem that is solved by the use of delayed sweep.

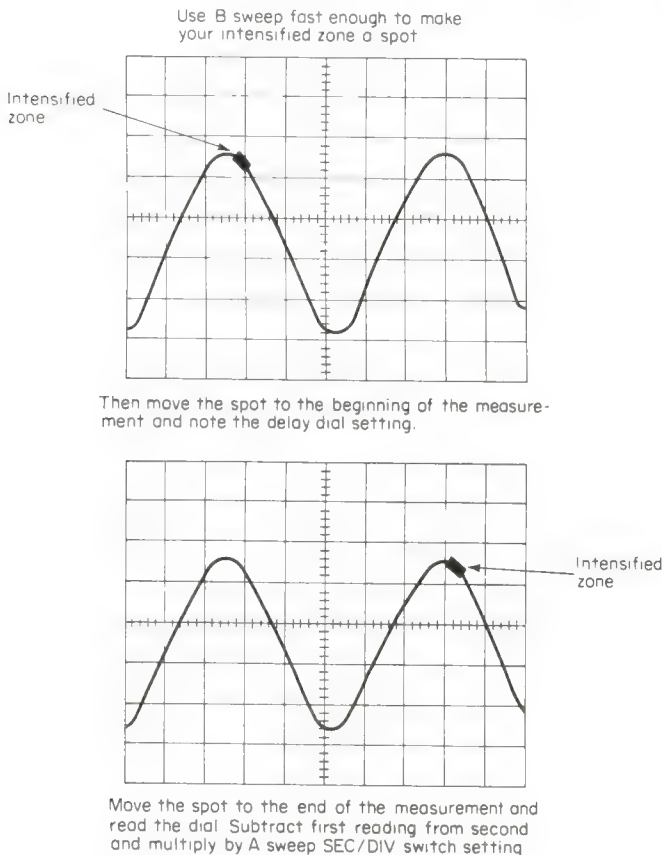
Another great measurement advantage of dual-time-base oscilloscopes is the separate B trigger. It eliminates jitter, which can prevent an accurate measurement any time you want to look at a signal that is not perfectly periodic. With two time bases and delayed sweep, you trigger the A sweep normally and move the intensified zone out to the portion of the waveform you want to measure. Then you set the oscilloscope up for a triggered B sweep, rather than letting the B sweep simply run after the delay time.

Besides examining signals at two different sweep speeds and seeing a jitterfree B sweep, you get increased timing measurement accuracy with a dual-time-base oscilloscope. On the front panel, the delay time control (labeled B DELAY TIME POSITION on the Tektronix 2215) is a measuring indicator as well as a B-sweep positioning device. Numbers in a window at the top of the dial are calibrated to the major divisions of the oscilloscope screen. The numbers around the circumference divide the major division into hundreds.

To make very accurate timing measurements (to  $\pm 1.5$  percent) with this control (see Fig. 24.29), do the following:

- Use the B-runs-after-delay horizontal operating mode.
- Place the intensified zone (or the B sweep waveform) where the timing measurement begins, and note the B DELAY TIME POSITION dial setting.
- Dial back to where the measurement ends, and note the reading there.
- Subtract the first reading from the second, and multiply by the A sweep SEC/DIV setting.

**example 24.7** You are using a delayed-sweep oscilloscope to measure the delay time between two gates in a digital circuit. With a SEC/DIV setting of 0.1 ms, you position the intensified zone



**Fig. 24.29** Time measurement using the calibrated delay dial.

at the rising edge of the first waveform and read the calibrated DELAY TIME MULTIPLIER dial as 3.14. Next you position the intensified zone at the rising edge of the second waveform. This time the dial reads 6.34. How much time has elapsed between the two events?

**solution** DELAY TIME dials on dual-time-base oscilloscopes are always calibrated in terms of the SEC/DIV settings. You find the difference between the two readings and multiply by the setting:  $(6.34 - 3.14) \times 0.1 \text{ ms} = 0.32 \text{ ms}$ .

**Single-Time-Base Oscilloscopes.** A few single-time-base oscilloscopes offer delayed-sweep measurements. Those that do may have measurement capabilities similar to those of the Tektronix 2213, which has three possible horizontal operating modes annotated on the front panel as NO DLY, INTENS, and DLY'D.

When you set the horizontal mode switch to NO DLY (no delay), only the normal sweep functions.

When you choose INTENS (intensified sweep), the oscilloscope will display the normal sweep, and the trace also will be intensified after a delay time. The amount of delay is determined by both the delay time switch (you can use 0.5  $\mu\text{s}$ , 10  $\mu\text{s}$ , or 0.2 ms) and the delay time multiplier control. The multiplier lets you pick from 1 to 20 times the switch setting.

You can use the intensified-after-delay mode to make sure the oscilloscope triggers where you want and then use the DLY'D (delayed) sweep mode to make the sweep start after the delay time chosen. After selecting this position, you can move the SEC/DIV control to a faster sweep speed and examine the waveform in greater detail.

## 24.4 EFFECTS OF INSTRUMENT PERFORMANCE

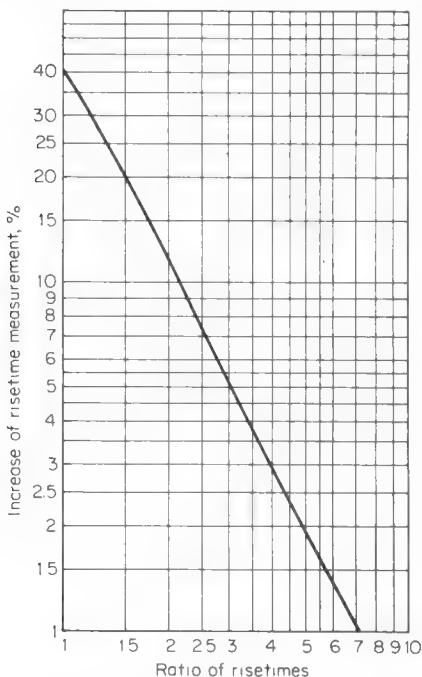
The nature of the world is such that any measurement affects the thing being measured. Most often the effect is negligible; in the case of oscilloscope measurements, the effects are also predictable.

There are two aspects to oscilloscope performance and its effects on your measurements: the design parameters of the instrument and its conformance to those parameters at the time of measurement. Making an instrument conform to its design parameters simply means proper calibration, including making sure the probe is properly compensated. But even with proper calibration, there will be some effect of the designed performance on your measurements.

**INSTRUMENT RISETIME AND MEASURED RISETIMES** The risetime of an oscilloscope is a very important specification because the measuring instrument's risetime affects the accuracy of your measured risetimes, as expressed by the following approximation:

$$T_{r(\text{meas})} = T_{r(\text{signal})}^2 + T_{r(\text{meas system})}^2$$

In practical terms, this means that the accuracy of a measured signal will be predictable and will depend on how much faster the oscilloscope is than the risetime you are measuring. If the measuring scope is 5 times faster than the observed signal, the measurement error can be as low as 2 percent. For measurement accuracies of 1 percent, it takes an oscilloscope 7 times faster, as you can see from Fig. 24.30. If the rise-



**Fig. 24.30** This chart enables you to calculate the percentage of risetime error.

times are equal, the error is a 41 percent increase.

**example 24.8** Measured risetimes depend on the risetime being measured and on the risetime of the measurement system. If the oscilloscope you are using has a risetime of 10 ns and you have measured a pulse risetime of 25 ns, what is the actual risetime?

**solution** Measured risetimes follow this formula:

$$\sqrt{T_{r(\text{meas})}} = \sqrt{T_{r(\text{signal})/2} + T_{r(\text{meas system})/2}}$$

Applying the values, you have

$$\begin{aligned} 25 &= \sqrt{T_{r(\text{signal})/2} + 10^2} \\ 25 &= \sqrt{T_r^2 + 100} \\ 625 &= T_r^2 + 100 \\ T_r^2 &= 525 \\ T_r &= 22.9 \text{ ns} \end{aligned}$$

**example 24.9** If the measuring oscilloscope has a 50-MHz bandwidth (BW), what will be the measured risetimes and percentage of error with signals having risetimes of 50, 35, 21, and 7 ns?

**solution** The scope risetime is 7 ns because

$$T_{r(\text{nanoseconds})} = \frac{350}{\text{BW (megahertz)}}$$

Then

$$T_{r(\text{meas})} = \sqrt{T_{r(\text{signal})/2} + T_{r(\text{meas system})/2}}$$

and

$$\begin{aligned} 50.5 &= \sqrt{50^2 + 7^2} \\ 35.7 &= \sqrt{35^2 + 7^2} \\ 22.1 &= \sqrt{21^2 + 7^2} \\ 9.9 &= \sqrt{7^2 + 7^2} \end{aligned}$$

The percentage of error, respectively, can be found by

$$\text{Measured risetime} - \text{signal risetime} / \text{Signal risetime} \times 100$$

or

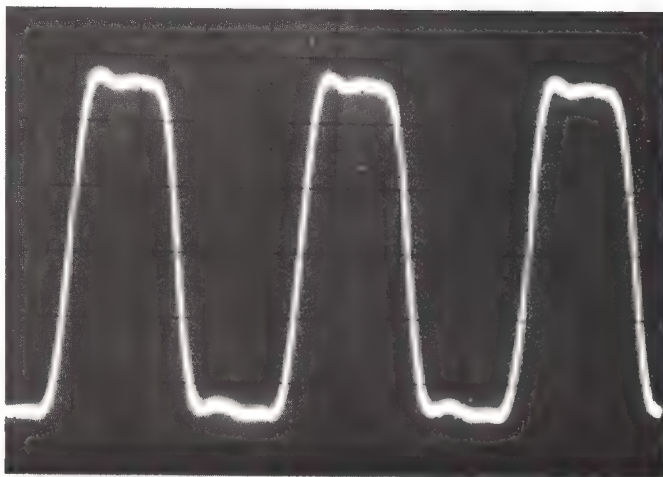
$$\begin{aligned} 1\% &= \frac{50.5 - 50}{50} \times 100 \\ 2\% &= \frac{35.7 - 35}{35} \times 100 \\ 5.4\% &= \frac{22.1 - 21}{21} \times 100 \\ 41.4\% &= \frac{9.9 - 7}{7} \times 100 \end{aligned}$$

**RELATING BANDWIDTH AND RISETIME** The vertical channels of an oscilloscope are designed for a broad bandpass, generally from some low frequency or direct current to a much higher frequency. This is the oscilloscope's bandwidth, which is specified by listing the frequency at which a sinusoidal input signal's amplitude has been attenuated to 0.707 of the amplitude of signals in the middle frequencies; this is called the  $-3\text{-dB}$  point.

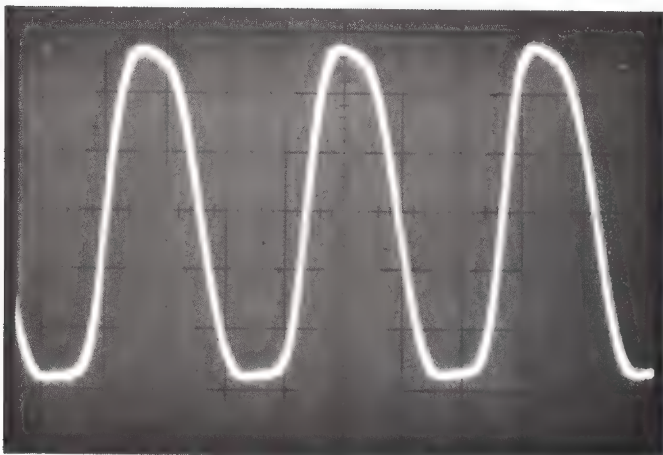
A bandwidth specification gives you an idea of the instrument's ability to handle high-frequency signals with a specified attenuation. But bandwidth specifications are derived from the instrument's ability to display sine waves. A 35-MHz oscilloscope will show a 35-MHz sine wave with only  $-3\text{-dB}$  attenuation, but the effects on a square wave at or near the oscilloscope's upper bandwidth limit will be much more severe because high-frequency information in the square wave will not be reproduced accurately. See Fig. 24.31 for an example.

Bandwidth specifications are based on the oscilloscope's ability to reproduce sine waves. The upper bandwidth is the frequency at which a sine wave is reduced to 0.707 of the amplitude shown at middle frequencies. Although this specification tells you how well the instrument reproduces sine waves, not every signal you examine is sinusoidal. Square waves, for example, have a great deal of high-frequency information in their rising and falling edges that will be lost as you approach the bandwidth limits of the instrument. To illustrate, the two CRT drawings in the figure show a 15-MHz square wave as reproduced by 60- and 35-MHz oscilloscopes.

The frequency response of most oscilloscopes is designed so that there is a constant that allows you to relate the bandwidth and risetime of the instrument. This constant is 0.35, and the risetime  $T_r$  and bandwidth BW are related by this approximation:



(a)



(b)

**Fig. 24.31** How a 15-MHz square wave is reproduced by (a) a 60-MHz oscilloscope and (b) a 35-MHz oscilloscope.

$$T_r = \frac{0.35}{BW}$$

This is a simple way to apply the formula:

$$T_{r(\text{nanoseconds})} = \frac{350}{BW (\text{megahertz})}$$

**example 24.10** What are the risetimes of 15-, 25-, 50-, and 100-MHz oscilloscopes?

**solution** Bandwidth and risetimes are related by

$$T_{r(\text{nanoseconds})} = \frac{350}{BW (\text{megahertz})}$$

Therefore,



Bandwidth, MHz	Risetime, ns
15	23.3
25	14
50	7
100	3.5

**MEASUREMENT SYSTEM BANDWIDTH** Like oscilloscopes, probes have a bandwidth; each has a specified range within which it does not attenuate the signal's amplitude more than -3 dB (0.707 of the original value). But do not assume that a 35-MHz probe and a 35-MHz oscilloscope give you a 35-MHz measurement capability. The combination will approximately equal the square root of the sum of the squares of the risetimes. In the 35-MHz examples, both the probe and the oscilloscope have risetimes of 10 ns:

$$T_{r(\text{system})} = T_{r(\text{oscilloscope})^2} + T_{r(\text{probe})^2}$$

$$T_r = 100 + 100 = 200 \text{ ns}$$

That works out to 14.14 ns, equivalent to a bandwidth of 24.75 MHz.

To get the full bandwidth from your oscilloscope, you need more bandwidth from the probe. Or you should use the particular probe designed for that instrument.

**example 24.11** If you have a 35-MHz oscilloscope and a 50-MHz probe, what are the bandwidth and risetime of your measurement system?

**solution** The oscilloscope will have a risetime of 10 ns, and the probe will have one of 7 ns because

$$T_{r(\text{nano seconds})} = \frac{350}{\text{BW (megahertz)}}$$

Scope:  $10 \text{ ns} = \frac{350}{35} \text{ MHz}$

Probe:  $7 \text{ ns} = \frac{350}{50} \text{ MHz}$

The risetime of the combination is then computed by

$$T_{r(\text{meas system})} = \sqrt{T_{r(\text{oscilloscope})^2} + T_{r(\text{probe})^2}}$$

$$T_r = \sqrt{10^2 + 7^2} = 12.2 \text{ ns}$$

Translating that to bandwidth yields 28.7 MHz.

**CIRCUIT LOADING** The probe designed for use with your oscilloscope is always the best choice, but even so there is an effect resulting from putting a probe in a circuit. This is called *circuit loading*. Circuit loading modifies the environment of the signals in the circuit you want to measure and changes the signals in the circuit under test. You minimize circuit loading when you use the correct probe and oscilloscope combination and when you have compensated the probe. But occasionally you must use another probe. In these cases, remember that circuit loading is resistive, capacitive, and inductive. For signals with low frequencies, the most important component of loading is resistance. To avoid significant circuit loading here, all you need is a probe with a resistance at least two orders of magnitude greater than the circuit impedance (100-M $\Omega$  probes for 1-M $\Omega$  sources, 1-M $\Omega$  probes for 10-k $\Omega$  sources, and so on).

When you are making measurements on a circuit that contains high-frequency signals, inductance and capacitance become important. You cannot avoid adding capacitance when you make connections, but you can avoid adding more capacitance than is necessary. One way to do that is to use an attenuator probe; its design greatly reduces loading. Instead of loading the circuit with capacitance from the probe tip plus the cable plus the oscilloscope's own input, the 10X attenuator probe introduces about 10 times less capacitance, as little as 10 to 14 pF. The only penalty is the reduction in signal amplitude from the 10:1 attenuation.

If you are not going to use the probes that came with your oscilloscope, also remember that proper termination is important to avoid unwanted reflections within the cable of the signal you want to measure. Probe-cable combinations designed to drive 1-M $\Omega$  inputs are engineered to suppress these reflections. But for 50- $\Omega$  oscilloscopes, 50- $\Omega$  probes should be



used. The proper termination is also necessary when you use a coaxial cable instead of a probe. If you use a 50- $\Omega$  cable and a 1-M $\Omega$  oscilloscope, be sure you also use a 50- $\Omega$  terminator at the oscilloscope input.

For all the reasons mentioned (probe bandwidth, loading, termination), the best way to ensure that the oscilloscope and probe measurement system has the least effect on your measurements is to use the probe recommended for your oscilloscope. And always make sure it is compensated.

## 24.5 STORAGE OSCILLOSCOPES

As useful as a conventional oscilloscope is, sometimes something more is needed. Maybe the signal you want to see happens so slowly that it is only a moving dot on the oscilloscope screen and you cannot see the waveshape. Maybe the signal is fast but happens only once in a while, too rarely for you to make a decent measurement. A storage oscilloscope could be the answer.

Conventional oscilloscopes allow you to evaluate events that are happening now; storage oscilloscopes let you save events and study them later. Because they store waveforms for later study, storage oscilloscopes also can be used to compare a waveform you saved earlier with signals occurring now. Additionally, storage oscilloscopes allow easy and accurate measurements of slowly changing phenomena on rapidly changing nonrepetitive waveforms whose images would otherwise flash across the CRT too quickly to be evaluated. Storage can also reduce the time it takes you to photograph waveforms because you can compose your pictures before taking out the oscilloscope camera.

Your measurement applications will determine which kind of storage oscilloscope you should use; read about CRT and digital storage oscilloscopes below to find the storage technology that provides the best instrument for your needs.

**CRT STORAGE OSCILLOSCOPES** There are two broad classes of storage oscilloscopes: CRT storage and digital storage. CRT storage oscilloscopes store the captured waveform within the CRT; digital oscilloscopes within a digital memory. In each category there are different technologies, and each has its own set of features and benefits.

**Bistable CRT storage** Bistable storage is the easiest CRT storage type to use. It is also the least expensive storage technology at present and will continue to be so until advances in digital storage change the picture. It features bright, long-lasting displays, but bistable storage displays have less contrast in comparison with other storage technologies.

This type of storage is named for the phosphor it uses. The phosphor coating the faceplate of a conventional oscilloscope has a single stable state, one in which the phosphor does not glow. When the electron beam strikes this phosphor, the phosphor glows for a short time afterward and then returns to its original state. The phosphor used in a bistable CRT storage oscilloscope, however, has two stable states: written and unwritten. Once this phosphor is struck by the electron beam, the phosphor will continue to glow; this allows waveforms to be displayed for as long as several hours or until they are erased by the operator.

The advantages of bistable storage make it particularly useful in applications involving signal comparisons, mechanical equipment measurements, and data recording.

In addition, most bistable storage oscilloscopes have split screens; you can control the top and bottom halves of the screen separately. This allows a reference waveform to be stored on one half the screen while you use the other half to see the effects of changes you are making on the circuit. The half of the screen that you are not using as a reference can be in either the storage or nonstorage mode.

**Variable-persistence CRT storage** If you do not need hours of stored waveforms and can use a storage oscilloscope that will hold a trace for a maximum of 30 to 60 min, variable-persistence CRT storage has advantages. In a variable-persistence CRT, the electron beam writes on a storage mesh first; then flood guns in the CRT paint waveforms on the screen where the storage mesh permits. Controls on the oscilloscope vary the charge on the mesh, allowing you to change the contrast between the trace and the background and to fine-tune how long the trace is stored.

The first capability provides easy viewing with high contrast between the dark background and bright waveforms, so variable-persistence storage provides the best displays for viewing traces with varying intensities (such as delaying and delayed sweeps or traces with Z axis intensity modulation).

Varying the length of time the stored waveform stays on the screen permits you to set up the oscilloscope so that the entire waveform can be viewed, yet the stored trace will fade from view just as a new waveform is being drawn. You could also view several new traces before the first one fades. Then you could see signal response variations as you make changes in the circuit.

You can also set up a variable-persistence storage oscilloscope to act as a display integrator; it can show a series of waveforms so that only the coincident portions are displayed. Any aberration or jitter not common to all traces will not be stored or displayed. A similar use is showing low-repetition-rate, fast-risetime signals that are not discernible on conventional CRTs; you can accomplish this by allowing each repetition to build up the trace brightness.

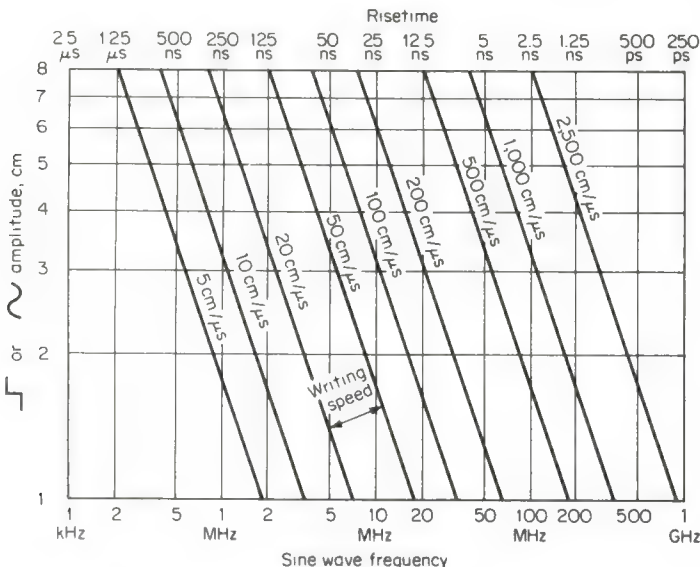
All these applications make variable-persistence CRT storage a good choice any time you do not need to keep waveforms on the screen for hours at a time.

**Fast-transfer CRT storage** Fast-transfer storage oscilloscopes use a CRT with a special intermediate mesh target designed to capture fast-moving sweeps of the electron beam. This target captures the waveform and then transfers it to another mesh, one optimized for long-term storage. As the name implies, the fast-transfer storage mode provides increased writing speed, letting you capture and store waveforms too fast for either bistable or variable-persistence oscilloscopes.

Since the second target can also be designed to offer either bistable or variable-persistence modes in combination with the transfer mesh or by itself, these oscilloscopes can be multi-mode storage instruments. With just the front panel controls, you can select the exact storage operating mode suited to your specific measurement situation.

**Stored writing speed** For CRT storage oscilloscopes, the most important specification (besides those common to all oscilloscopes) is the stored writing speed. This figure is expressed as either centimeters per unit of time or graticule divisions per unit of time. Often divisions per microsecond are more meaningful in terms of your measurements because some oscilloscopes have nonstandard sized graticules (i.e., other than 1-cm<sup>2</sup> major divisions). However, centimeters per microsecond is an absolute measurement which allows you to compare different families of storage technologies.

In either case, the specification is dependent on the speed of the input signal and on the size of the trace that the oscilloscope must draw. If you know the pulse risetime or sine wave frequency of the input signal and the size of the waveform you want displayed, you can use the chart of Fig. 24.32 to determine which writing speed you need.



**Fig. 24.32** Chart to tell whether a CRT storage oscilloscope can display the desired signals.

The stored writing speed is the specification to use to determine whether a CRT storage oscilloscope can display the signals you must see. This specification depends on the phosphor and the electronics of the CRT as well as the input signal. To use the chart, start with either frequency (for sine waves) or risetime (for step signals) and follow the vertical axis to the point where it intersects the size of the display you want. Use the diagonal writing speed line to the left of the intersection to find the specification you need for your measurements.

**DIGITAL STORAGE OSCILLOSCOPES** Digital storage oscilloscopes are easy to use and give you crisp, clear displays. Because the data are stored in a digital memory, no fading or blooming of the trace on the CRT phosphor will occur, and storage time is essentially unlimited. Because the waveforms you capture are stored as 1s and 0s in a digital memory and not in a CRT that will eventually have to be replaced, the life-cycle cost of a digital storage oscilloscope is likely to be lower.

Also you can do some things with digital storage that you cannot do with any other kind of storage, such as

- *Pretrigger viewing.* Because the trigger of a digital oscilloscope can be used to stop (as well as start) data acquisition, you can actually capture events that happened before the trigger.
- *Babysitting.* Because of its triggering and data acquisition techniques, you can use a digital storage oscilloscope in an untended mode to trap and save events that happen during the night or any other time when there is no operator.
- *Record keeping.* You can attach a digital storage oscilloscope to a controller and dump the waveforms it captures into a computer memory for permanent storage.
- *Signal processing.* Once the data are stored digitally, you can manipulate them, performing calculations such as rms voltage, averaging a number of samples of the input signal to reduce the effects of noise, etc.
- *Glitch catching.* Some digital storage oscilloscopes offer the capability of recording signals at what amounts to more than one sweep speed. You can look at signals over long periods yet catch quick transient events.

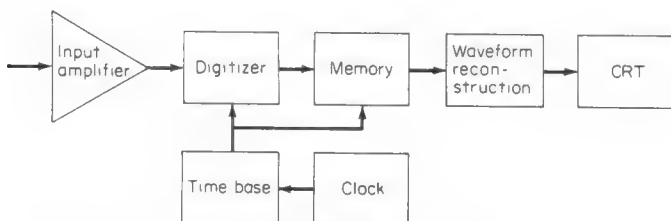
The Tektronix 468 portable scope shown in Fig. 24.39 (at the end of this chapter) is one example of a digital storage oscilloscope with all the features mentioned above.

**How digital storage works** Because digital storage is a relatively new concept for oscilloscopes, we take a brief look at how it works. Start with the block diagram in Fig. 24.33. The front end of a digital storage oscilloscope is like one on a conventional oscilloscope or a CRT storage oscilloscope in that input signals are conditioned and coupled by a vertical system. Then the differences start. The analog input signal is digitized by an analog-to-digital converter (ADC) and stored in a memory. The rate at which the signal is digitized and stored is timed by the oscilloscope's digital clock and controlled with your time-base control selections. Then the stored data are converted back to analog form by a waveform reconstruction technique and displayed on the oscilloscope's CRT.

Within digital storage oscilloscopes there are two main techniques of digitizing signals: real-time sampling and equivalent-time sampling. Which technique is used has a direct effect on the applications of the instruments. Only real-time sampling can capture single-shot signals, while both can capture repetitive signals.

In real-time sampling, the oscilloscope digitizes the input waveform sequentially at equal intervals of time during a single sweep.

In equivalent-time sampling, two methods are used. One, random (or asynchronous) sampling, takes a sample of a different part of the input waveforms each time. The other



**Fig. 24.33** Block diagram of a digital storage oscilloscope.



method, sequential equivalent-time sampling, samples the input waveforms a little later in the waveform after each trigger. In both cases, eventually a complete picture of the input signal is built during the repetitions.

The advantage of real-time sampling is that it can capture single-shot signals. The advantage of equivalent-time sampling is that it can capture signals that are 10 times or more faster than those that real-time sampling can handle.

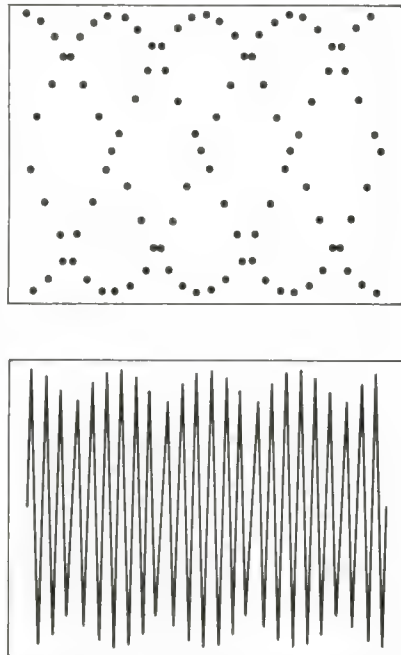
**Comparing digital and CRT storage** Along with the many advantages of digital storage oscilloscopes over CRT storage oscilloscopes, there are some drawbacks. One is the nature of the errors you can get with digital storage. For example, if you are trying to look at a signal that is faster than the bandwidth specification of a conventional or CRT storage oscilloscope, the amplitude of the displayed signal is reduced. However, digital storage oscilloscopes have what is called a *Nyquist frequency*, which is one-half the digitizing rate of the oscilloscope. If you try to measure a signal at or above the Nyquist limit, you get *aliasing*. An aliased signal does not look wrong (the amplitude is not reduced as it would be by an analog oscilloscope), and if you do not know it happened, your measurements will be unreliable. See Fig. 24.34 for an example.

It is possible to misinterpret a digital storage oscilloscope display when "perceptual aliasing" occurs; it is a sort of optical illusion that occurs with dot displays. For example, the original sine wave input signal is not apparent in the first drawing; but when vectors are added to connect the data points, the sine wave is easier to see. In a metaphorical sense, the aliasing of an input signal greater than the Nyquist limit of the oscilloscope is similar. Then the samples taken beyond the storage bandwidth limit result in a display that looks as if the input signal were lower than the bandwidth limit. Unfortunately in this case, the data in the memory are wrong, the frequency information is lost, and adding vectors cannot correct the mistake.

Another difference between CRT and digital storage involves bandwidth. In analog oscilloscopes, the bandwidth of the instrument never changes. But the storage bandwidth of a digital oscilloscope is not always the same, because when you change time-base settings, you also change the digitizing rate (unless the oscilloscope has a feature like the envelope mode of the Tektronix 468 digital storage oscilloscope).

Last, there is resolution. Resolution is the ability to discriminate between items. On analog oscilloscope screens, the resolution is continuous. It depends on how sharp a trace the electron beam can draw on the screen and how carefully you make the measurement; but you can achieve great resolution. For digital oscilloscopes, however, the resolution is dependent on the hardware. The vertical resolution you need to discriminate between small voltage differences depends on the resolution of the analog-to-digital converter. The horizontal (time) resolution is dependent on how many data words are available in the oscilloscope's memory. See Table 24.2 to translate bits of digital storage oscilloscope resolution to percentages.

Use the columns in Table 24.2 to relate digital storage oscilloscope resolution specifications to percentages. For bits of vertical resolution, use the first column. To determine horizontal resolution in percentages, find the memory size used to store one waveform in the middle column: if 128 data words are used to store a waveform, you will be able to resolve 1 in 128 parts, the equivalent of 0.8 percent. Remember, though, that the resolution may not be the



**Fig. 24.34** An example of "perpetual aliasing" on a digital storage oscilloscope.

**TABLE 24.2** Relation of Digital Storage Oscilloscope Resolution to Percentages

Bits	Parts	Percentage
1	2	50.0
2	4	25.0
3	8	12.5
4	16	6.25
5	32	3.125
6	64	1.563
7	128	0.781
8	256	0.391
9	512	0.195
10	1,024	0.098
11	2,048	0.049
12	4,096	0.024
13	8,192	0.012
14	16,384	0.006

limiting specification; for example, a digital storage oscilloscope could have an accuracy of 3 percent in the analog amplifiers of its vertical system. In that case, accuracy will be 3 percent, even if the instrument has 15 bits of vertical resolution.

**Digital storage oscilloscope specifications** For real-time sampling digital storage oscilloscopes, there is a useful storage bandwidth specification that expresses the highest frequency sine wave that can be captured in a single sweep and displayed so that you can make measurements. Both the digitizing rate (how often the oscilloscope takes samples) and the display reconstruction technique (how the oscilloscope displays what is in its memory) must be taken into account in the useful storage bandwidth (see Fig. 24.35).

Figure 24.35 illustrates the effects of reconstruction techniques on the usefulness of a digital storage oscilloscope display. A 10-MHz sine wave was digitized at 25 MHz, and each sample is displayed on the screen. In the first case, only dots are used to indicate the data points, and the sine wave is not readily apparent. In the second drawing, vectors have been generated to join each data point, and the waveshape is more easily seen. And in the last drawing, the vectors have been smoothed by a sine interpolator, and the sine wave is obvious. Had the input signal been a step function—as opposed to a sine wave—a vector display would be more accurate than sine interpolation.

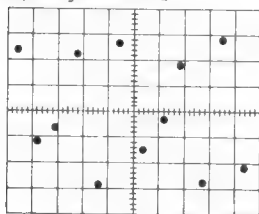
For digital oscilloscopes using equivalent-time sampling, the specification is equivalent-time bandwidth; this is the highest frequency signal that can be stored and displayed with less than 3-dB signal amplitude loss.

Other specifications of interest to digital oscilloscope users—besides the specifications, such as accuracy and linearity, that are common to all oscilloscopes—are as follows:

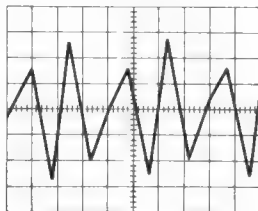
**Maximum digitizing rate**, or how often the instrument takes samples of the input signal.

**Vertical resolution** (usually expressed in bits of resolution) describes how finely the instrument can discriminate between signals very much alike in voltage.

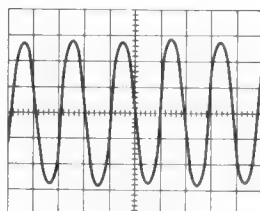
Input signal 10 MHz



(a)



(b)



(c)

**Fig. 24.35** Effects of reconstruction techniques with a digital storage oscilloscope. (a) Dot display. (b) Pulse interpolator. (c) Sine interpolator. Digitizing rate is 25 MHz.

*Data words per waveform.* Horizontal resolution depends on how many words of digital memory are used to store the captured waveforms.

## 24.6 SELECTING AN OSCILLOSCOPE

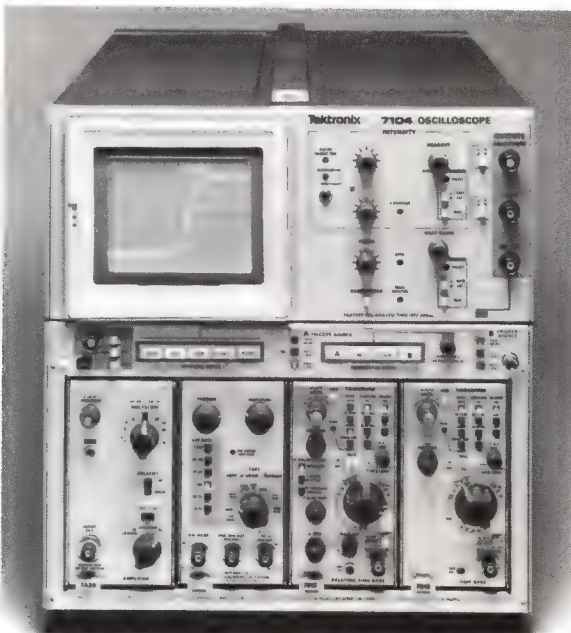
Picking a test and measurement instrument involves three parameters: suitability, service, and support. Suitability pertains to the instrument itself, while service and support are attributes of the manufacturer. The last two parameters are easier to evaluate. For service, simply find out how many service centers there are, where they are located, what kinds of service contracts are available, and whether the centers service the brand you want to buy exclusively (a plus). To determine how much support you can expect, find out whether the instrument manuals are usable, whether there are application notes and collateral literature available, and whether customer operation and service training are possible. Even if you do not take advantage of all the customer support, you can have confidence in the manufacturer when it is there.

Suitability is another matter. You know the applications best, but the manufacturer's sales engineers know the instruments best. Investigate what is available, read the literature, and then work with your local sales engineer to make the best decision. Figures 24.36 through 24.39 illustrate the broad range of oscilloscopes available today.

The Tektronix 7104 general-purpose oscilloscope is an example of the plug-in oscilloscope concept (Fig. 24.36). By adding time bases and vertical amplifiers to a mainframe, you can build the exact oscilloscope necessary for your applications. The configuration pictured will give you state-of-the-art capabilities: 1-GHz bandwidth, 200-ps sweep speeds, and a horizontal bandwidth of 350 MHz for XY measurements.

The Tektronix 2215 is an example of a portable scope (Fig. 24.37). It features dual 60-MHz channels, two time bases (for delayed-sweep timing measurements accurate to  $\pm 1.5$  percent), light weight (13.5 lb), and a fastest sweep of 5 ns.

The Tektronix 2336 is one of three 100-MHz oscilloscopes designed for field service in harsh environments (Fig. 24.38). These instruments are rugged and compact and have an



**Fig. 24.36** Example of the plug-in oscilloscope concept. (Courtesy Tektronix.)





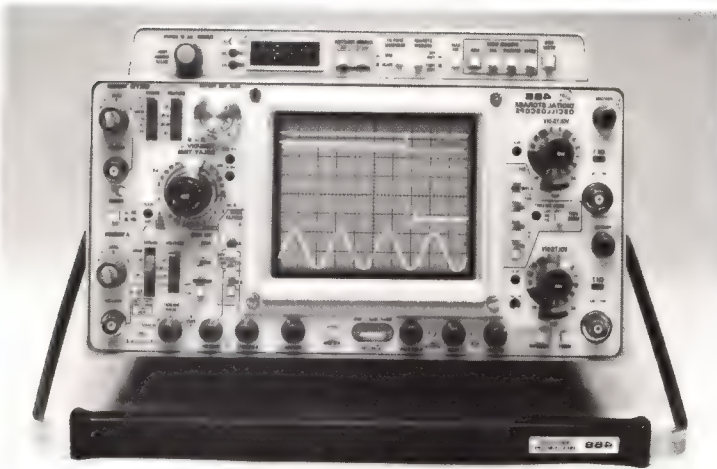
**Fig. 24.37** Example of a portable oscilloscope. (Courtesy Tektronix.)

outstanding resistance to electromagnetic interference. A liquid crystal display on the 2336 gives you a digital readout of delta time measurements; the 2337 adds an LCD readout of digital multimeter measurements.

The Tektronix 468 digital storage oscilloscope offers the best of two worlds (Fig. 24.39). Used as a normal real-time oscilloscope, it is a 100-MHz portable instrument. But for applications where signals must be captured now and viewed later, the 468 is a 10-MHz storage



**Fig. 24.38** A 100-MHz oscilloscope designed for field service in harsh environments. (Courtesy Tektronix.)



**Fig. 24.39** Example of a digital storage oscilloscope. (Courtesy Tektronix.)

oscilloscope. Best of all, it is a digital storage oscilloscope, which means that the waveforms are recorded as digital data in a memory, giving users more measurement opportunities than any analog storage oscilloscope.

## 24.7 BIBLIOGRAPHY

- Douglas-Young, John: *Practical Oscilloscope Handbook*, Parker Publishing Co., West Nyack, N.Y., 1979.
- Lenk, John, D.: *Handbook of Oscilloscopes: Theory and Application*, Prentice-Hall, Englewood Cliffs, N.J., 1982.
- Roth, Charles H., Jr.: *Use of the Dual Trace Oscilloscope: A Programmed Text*, Prentice-Hall, Englewood Cliffs, N.J., 1982.



# INDEX

- Acceptance cone, **21-6**
- Access time, **15-18**
- Accumulator, **18-7**
- Accumulator addressing, **18-24**
- Accuracy, instrumental, **7-3, 7-22**
- Active element, **6-1**
- Active filters, **11-26 to 11-29, 22-1 to 22-22**
  - advantages of, **22-1**
  - amplifying, **22-3**
  - attenuation rate of, **22-14**
  - band-reject, **11-14, 11-29**
  - bandpass, **11-12 to 11-14, 11-29, 22-17**
  - bandstop, **11-14, 22-20**
  - Bessel, **22-5**
  - broadband, **22-16**
  - Butterworth, **11-22, 11-26 to 11-28, 22-5**
  - cascaded, **22-13**
  - Chebyshev, **11-22, 22-5**
  - circuit quality factor of, **22-5**
  - circuitry of, **22-6 to 22-9, 22-12**
  - design data for, **22-9 to 22-12**
  - first-order, **22-2**
  - high-pass (*see* High-pass filters)
  - low-pass (*see* Low-pass filters)
  - narrowband, **22-17**
  - second-order (*see* Second-order filters)
  - third-order, **22-12 to 22-16**
- Active region, **14-5**
- A/D (analog-to-digital) conversion, **15-23**
- Address buffer, **18-8**
- Address latch, **18-8**
- Addressing, **18-11, 18-24**
  - direct, **18-11, 18-12**
  - extended, **18-23**
  - immediate, **18-11, 18-12**
- Addressing (*Cont.*):
  - implied, **18-24**
  - indexed, **18-11, 18-24**
  - indirect, **18-11, 18-12**
  - relative, **18-24**
- Admittance parameters, **12-15**
- Air-core coils, **3-3 to 3-5, 3-7 to 3-9**
- Air-depolarized batteries, **17-29**
- Air-variable capacitors, **2-11**
- Air-zinc batteries, **17-30**
- Alkaline batteries, **17-11 to 17-13**
- ALU (arithmetic logic unit), **18-8**
- Aluminum electrolytic capacitors, **2-9**
- Ambient temperature, **2-4**
- Ammeters, **7-3**
- Ampere-turn (unit), **4-9, 4-11**
- Ampere's law, **7-9**
- Amplifiers, **12-1 to 12-28**
  - BJT, **12-11 to 12-15**
  - cascaded, **12-21, 13-12**
  - common-base, **8-11, 12-6, 12-13**
  - common-collector, **12-7, 12-13**
  - common-emitter (*see* Common-emitter amplifiers)
  - common-source, **12-16, 12-20**
  - difference, **7-17, 13-3**
  - differential, **13-10**
  - feedback, **12-26 to 12-28**
  - FET, **12-15 to 12-17**
  - frequency response of (*see* Frequency response)
  - inverting, **13-2**
  - JFET, **12-9**
  - logarithmic, **13-5**
  - microwave, **20-29**

Amplifiers (*Cont.*):

- MOSFET (*see* MOSFET amplifiers)
- noninverting, 13-3
- operational (*see* Operational amplifiers)
- power (*see* Power amplifiers)
- stabilization of, (*see* Stabilization)
- summing, 13-4
- terminology for describing, 12-2 to 12-4
- totem-pole, 13-13, 15-4
- transimpedance, 21-32
- voltage, 21-32

## Amplifying filters, 22-3

## Analog-to-digital (A/D) conversion, 15-23

## AND gate, 14-7

## Antennas, 20-33 to 20-38

- arrays of, 20-36
- beamwidth of, 20-35
- directivity of, 20-34, 20-35
- effective area of, 20-35
- focusing, 20-40
- gain of, 20-34
- horn, 20-40
- isotropic, 20-33
- microwave, 20-38 to 20-40
- pattern multiplication of, 20-37
- patterns of, 20-33
- polarized, 20-34
- reciprocal, 20-33
- spiral, 20-38

## Archimedes spiral antennas, 20-38

## Arithmetic instructions, 18-16

## Arithmetic logic unit (ALU), 18-8

## Aspect ratio, 9-12

## Astable multivibrators, 14-17

## Asynchronous counter, 15-14

## Asynchronous sampling, 19-20, 19-27

## Attenuation in optical fibers, 21-7 to 21-12

## Attenuation rate, 22-14

## Attenuators, 20-22, 20-29

## variable, 20-22

## Automatic-zero circuit, 23-2

## Autotransformers, 5-21 to 5-26

## switching of, 5-26

## variable, 5-24 to 5-26

## Avalanche breakdown, 8-8

## Avalanche photodiodes, 21-32

## Avalanche-type diodes, 8-29

## Background concentration, 9-7

## Band-reject filters, 11-14, 11-29

## Bandpass filters, 11-12 to 11-14, 11-29, 22-17

## Bandstop filters, 11-14, 22-20

## Bandwidth:

- amplifier, 12-3, 12-28
- of optical fibers, 21-12 to 21-14
- oscilloscope, 24-15, 24-25, 24-27
- of tuned circuits, 10-4, 10-15

## Barkhausen criterion, 12-29

## Base diffusion, 9-8

## Base-emitter junction, 8-9

## Base-emitter resistance, 12-17

## Base masks, 9-8

## Base-spreading resistance, 12-17

## Batteries, 17-1 to 17-33

## air-depolarized, 17-29

## air-zinc, 17-30

## alkaline, 17-11 to 17-13

## charging of, 17-24, 17-26

## cuprous chloride-magnesium, 17-28

dry-cell (*see* Primary batteries)

## fuel-cell, 17-31 to 17-33

## lead-acid, 17-17 to 17-19, 17-24

## LeClanche, 17-5 to 17-11

## magnesium dry, 17-15, 17-28

## mercury dry, 17-13 to 17-15

## nickel-cadmium, 17-19 to 17-22, 17-26

## organic depolarized, 17-28

primary (*see* Primary batteries)rechargeable (*see* Secondary batteries)

## reserve, 17-3, 17-29

secondary (*see* Secondary batteries)

## silver chloride-magnesium, 17-29

## silver zinc, 17-22 to 17-24, 17-26

## Beam-lead device, 8-30

## Beam-lead isolation, 9-9

## Beamwidth, 20-35

## Bessel filters, 22-5

## Biasing:

## common-base amplifier, 12-6

## common-collector amplifier, 12-6

## common-emitter amplifier, 12-4 to 12-6

## depletion-type MOSFET amplifier, 12-9 to 12-11

## enhancement-type MOSFET amplifier, 12-11

## JFET amplifier, 12-9

## Bidirectional bus, 18-3

## Bidirectional triode thyristors, 8-21

## Bifilar resistor winding, 1-10

## Binary display mode, 19-9, 19-17, 19-27

## Binary numbers, 14-1 to 14-3

## Binary-to-decimal conversion, 14-3

## Bipolar junction transistors (BJT), 8-9 to 8-13

## amplifying, 12-11 to 12-15

## diffused, 9-10

- Bipolar junction transistors (BJT) (*Cont.*):
  - integrated circuits and, 9-10
  - operation of, 8-10 to 8-12
  - parameters for describing, 8-12
- Bistable cathode-ray tube (CRT)
  - oscilloscopes, 24-28 to 24-30
- Bistable multivibrators, 14-15
- Bit, 18-1
- BJT (*see* Bipolar junction transistors)
- Blanking, 19-18, 19-27
- Bode plot, 13-7
- Branch control instructions, 18-18
- Break frequency, 13-7
- Breakdown, avalanche, 8-8
- Breakdown voltage, 2-5
- Bridge rectifiers, 16-7
- Broadband filters, 22-16
- Bulk-film fixed resistors, 1-9
- Buried crossover, 9-15
- Bus, 18-3
- Butterworth filters, 11-22, 11-26 to 11-28, 22-5
- Bypass capacitors, 12-7, 12-9
- Byte, 18-1
- Cables:
  - coaxial, 20-4
  - optical fiber, 21-14 to 21-17
- Calorimetric method, 7-8
- Capacitance, 2-4
  - depletion, 12-18
  - diffusion, 12-18
  - distributed, 3-11
  - Miller, 12-18
  - temperature coefficient of, 2-5
  - temperature variation of, 2-16
  - total diode, 2-16
- Capacitive reactance, 2-5, 10-1, 10-6
- Capacitors, 2-1 to 2-17
  - air-variable, 2-11
  - aluminum electrolytic, 2-9
  - applications of, 2-1
  - bypass, 12-7, 12-9
  - ceramic chip, 2-15
  - ceramic fixed, 2-8
  - charge on, 2-3
  - chip, 2-13 to 2-15
  - color coding of, 2-11
  - coupling, 12-5
  - decoupling, 12-9
  - electrolytic, 2-9
  - energy-storage, 2-9
  - energy stored in, 2-4
- Capacitors (*Cont.*):
  - fixed, 2-7 to 2-11
  - glass, 2-8
  - impedance in, 2-6
  - integrated circuits and, 9-12
  - junction, 9-12
  - metallized dielectric, 2-9
  - mica, 2-7 to 2-8
  - paper, 2-8
  - parallel, 2-10
  - plastic film, 2-8
  - series, 2-10
  - thin-film, 9-12
  - trimmer, 2-12
  - vacuum, 2-9
  - vacuum-variable, 2-11
  - variable, 2-11 to 2-13
  - voltage-variable, 2-15 to 2-17
  - volumetric efficiency of, 2-7
- Carbon-composition fixed resistors, 1-8
- Carbon-film fixed resistors, 1-9
- Carbon variable resistors, 1-15
- Cascaded amplifiers, 12-21, 13-12
- Cascaded filters, 22-13, 22-15 to 22-16
- Cathode-ray tube (CRT) oscilloscopes, 24-28 to 24-30
- CCD (charge-coupled device), 8-28
- Center frequency, 10-18, 11-5
- Center triggering, 19-16, 19-26
- Ceramic chip capacitors, 2-15
- Ceramic filters, 11-23
- Ceramic fixed capacitors, 2-8
- Cermet fixed resistors, 1-9
- Cermet variable resistors, 1-15
- Characteristic impedance:
  - of filters, 11-9
  - of transmission lines, 20-3, 20-8
- Charge-coupled device (CCD), 8-28
- Charging of batteries, 17-24, 17-26
  - nickel-cadmium, 17-26
  - silver-zinc, 17-26
- Chebyshev filters, 11-22, 22-5
- Chip capacitors, 2-13 to 2-15
- Chip resistors, 1-16
- Circuit analysis:
  - conductance and, 6-8
  - energy and, 6-5
  - Kirchhoff's laws and, 6-5
  - laws of, 6-5 to 6-8
  - notation for, 6-1
  - periodic waveforms and (*see* Periodic waveforms)
  - power and, 6-5
  - of RC circuits, 6-3



Circuit analysis (*Cont.*):

- of *RL* circuits, 6-3
- terminology for describing, 6-1 to 6-3

## Circuit integration, 9-2 to 9-9, 9-13 to 9-16

- artwork, 9-4
- base diffusion, 9-8
- base mask, 9-8
- buried crossover, 9-15
- collector mask, 9-6
- contact mask, 9-8, 9-14
- diffusion, 9-6
- emitter, 9-8
- epitaxial layer, 9-3
- etching, 9-7
- interconnection mask, 9-9, 9-14
- isolation mask, 9-14
- metallization, 9-8
- n+ diffusion, 9-8, 9-14
- oxidation, 9-8
- oxide layer, 9-3
- p-type diffusion, 9-7, 9-14
- p-type substrate, 9-3
- photoresist, 9-4
- resist, 9-8
- (*See also* Integrated circuits)

## Circuit loading, 24-27

## Circuits:

- digital (*see* Digital circuits)
- integrated (*see* Integrated circuits)
- logic (*see* Logic circuits)
- magnetic (*see* Magnetic circuits)
- tuned (*see* Tuned circuits)

## Circulating registers, 15-16

## Circulators, 20-26

## Clamping diodes, 15-5

## Clock in data acquisition function of logic analyzer, 19-6, 19-10

## Clock qualifiers, 19-12, 19-17

## Clock width, 19-27

## Clocked R-S (reset-set) flipflop, 15-9

## Closed-loop gain, 12-27, 13-7

## Closed-loop regulators, 16-17

## CMOS (complementary metal-oxide semiconductor) logic, 15-7

## Coaxial cables, 20-4, 20-6

## Coefficient of coupling, 3-10

## Coefficient of reflection, 20-13

## Coils, 3-1

- air-core, 3-3 to 3-5, 3-7 to 3-9
- coefficient of coupling between, 3-10
- dielectric loss in, 3-10
- distributed capacitance in, 3-11
- eddy currents in, 3-10
- ferrous core (*see* Ferrous core coils)

Coils (*Cont.*):

- left-hand rule for, 4-6
- mutual inductance in (*see* Mutual inductance)
- nonferrous core, 3-3
- in parallel, 3-6
- resistance of, at radio frequencies, 3-10
- right-hand rule, 4-6
- in series, 3-6
- skin effect in, 3-10

## Collector-base junctions, 8-9

## Collector characteristics, common-emitter, 8-13

## Collector dissipation, 12-23

## Collector-emitter resistance, 12-18

## Collector masks, 9-6

## Color coding:

- of capacitors, 2-11
- of resistors, 1-11

## Colpitts oscillators, 12-30

## Common-base amplifiers, 8-11, 12-6, 12-13

## Common-collector amplifiers, 12-7, 12-13

## Common-emitter amplifiers, 8-11

- biasing, 12-4 to 12-6
- figures of merit of, 12-19
- hybrid model of, 12-13
- hybrid-pi model of, 12-17 to 12-19
- low-frequency response of, 12-21
- stabilizing, 12-7 to 12-9

## Common-emitter collector characteristics, 8-13

## Common-mode rejection, 21-13

## Common-mode rejection ratio, 13-11

## Common-mode signal, 13-10

## Common-mode voltage range, 13-18

## Common-source amplifiers, 12-16, 12-20

## Comparators, 15-23

## Compare display mode, 19-18, 19-27

## Compass, magnetic, 4-3

## Complementary metal-oxide semiconductor (CMOS) logic, 15-7

## Complementary symmetry amplifiers, 12-25

## Conductance, 6-8

## Conductive plastic variable resistors, 1-15

## Conductors:

- eddy currents in, 3-10
- resistance of, at radio frequencies, 3-10

## Connecting mismatch loss, optical fiber, 21-20

## Connectors, optical fiber, 21-18

Constant-*k* filters, 11-7, 11-9 to 11-15

- band-reject, 11-14
- bandpass, 11-12 to 11-14
- design of, 11-9

- Constant- $k$  filters, (*Cont.*):
  - high-pass, 11-11
  - low-pass, 11-9 to 11-11
- Contact mask, 9-8, 9-14
- Contact resistance, 1-6
- Continuity, electric, 7-5
- Conversion:
  - analog-to-digital (A/D), 15-23
  - binary-to-decimal, 14-3
  - decimal-to-binary, 14-2
  - digital-to-analog (D/A), 15-20 to 15-23
- Conversion efficiency, 2-17
- Coordinatograph, 9-5
- Copper loss in transformers, 5-15
- Core-loss current, 5-9
- Core-loss test, transformers, 5-16
- Coulomb's law, 4-4
- Counters, 15-13 to 15-15
  - asynchronous, 15-14
  - down, 15-14
  - modulo, 15-14
  - ripple, 15-13
  - synchronous, 15-14
  - up, 15-14
  - up-down, 15-14
- Couplers, 20-24
  - optical fiber, 21-19
- Coupling capacitors, 12-5
- Coupling coefficient, 3-10
- Coupling transformers, 12-23
- Crest factor, 23-6
- Critical angle, 21-6
- Critical coefficient of coupling, 10-21
- Critical resistance, 1-4
- Crossover, buried, 9-15
- Crossover distortion, 12-25
- Crossover frequency, 13-7
- CRT (cathode-ray-tube) oscilloscopes, 24-28
  - to 24-30
- Crystal filters, 11-22
- Crystal oscillators, 12-31
- $C_r$  ratio, 2-16
- Cuprous chloride-magnesium batteries, 17-28
- Current:
  - core-loss, 5-9
  - dark, 8-28
  - eddy, 3-10
  - load, 5-11
  - magnetizing, 5-9
  - meter, 7-13
  - no-load, 5-8
  - peak, 8-23
  - source of, 6-2
- Current (*Cont.*):
  - valley, 8-23
- Current-controlled current source, 6-2
- Current-controlled voltage source, 6-2
- Current division, 12-15
- Current feedback, 12-28
- Current-feedback stabilization, 12-8
- Current gain, 12-3
- Current-reading meters, 7-13
- Current sink, 13-11
- Current-to-voltage converters, 13-6
- Current transformers, 5-27
- Cutoff frequency, 2-16, 11-7
- Cutoff region, 14-5
- Cycle time, 15-18
- D/A (digital-to-analog) conversion, 15-20 to 15-23
- Damping factor, 22-5
- Dark current, 8-28
- Darlington pairs, 13-17, 15-4
- Data acquisition, logic analyzer, 19-8, 19-10
  - to 19-12
    - asynchronous sampling, 19-20, 19-27
    - clock, 19-6, 19-10
    - clock qualifier, 19-12, 19-27
    - glitch capturing mode, 19-21, 19-27
    - hold time, 19-11, 19-27
    - input, 19-10
    - setup time, 19-11, 19-27
    - start function, 19-12
    - synchronous sampling, 19-20, 19-27
    - terminology for describing, 19-27
- Data field, 19-27
- Data transfer instructions, 18-13
- DC (*see* Direct-current *entries*)
- Decibel (unit), 20-2
- Decimal-to-binary conversion, 14-2
- Decoders, 18-22
- Decoupling capacitors, 12-9
- Delay flipflop, 15-12
- Delay time, 14-18
- Delayed triggering, 19-15, 19-26
- Depletion capacitance, 12-18
- Depletion-type MOSFET, 8-15, 12-9 to 12-11
- Derating curve, 1-3
- Detectors, microwave, 20-26
- Diamagnetism, 4-3
- Dielectric, 2-1
- Dielectric constant, relative, 2-3
- Dielectric isolation, 9-9
- Dielectric loss, 3-10 to 3-13

- Dielectric transmission lines, **20-13**
- Difference amplifiers, **7-17, 13-3**
- Difference-mode signal, **13-10**
- Differential amplifiers, **13-10**
- Diffused bipolar junction transistor (BJT), **9-10**
- Diffusion:
  - base, **9-8**
  - in circuit integration, **9-6**
  - n+, **9-8, 9-14**
  - p-type, **9-7, 9-14**
- Diffusion capacitance, **12-18**
- Digital circuits, **14-1 to 14-19**
  - binary numbers and, **14-1 to 14-3**
  - binary-to-decimal conversion and, **14-3**
  - decimal-to-binary conversion and, **14-2**
  - diode switches in, **14-4, 14-17**
  - integrated (*see* Digital integrated circuits)
  - logic (*see* Logic circuits)
  - multivibrators and (*see* Multivibrators)
  - transistor switches in, **14-4, 14-18**
- Digital displays, **7-19**
- Digital integrated circuits, **15-1 to 15-23**
  - analog-to-digital converters in, **15-23**
  - CMOS logic in, **15-7**
  - counters in, **15-13 to 15-15**
  - digital-to-analog converters in, **15-20 to 15-23**
  - ECL in, **15-6**
  - flipflop (*see* Flipflops)
  - semiconductor memories and (*see* Semiconductor memories)
  - shift registers in, **15-15**
  - terminology for describing, **15-1**
  - transistor-transistor logic (*see* TTL circuits)
- Digital multimeters, **7-6, 7-18 to 7-22, 23-1 to 23-10**
  - accuracy of, **7-22**
  - alternating-current conversion in, **23-4, 23-7 to 23-10**
  - direct-current input amplification, **23-2**
  - direct-current input attenuation, **23-2**
  - display technology, **7-19**
  - dual-slope, **23-7**
  - guarded, **23-12**
  - integrating, **23-7 to 23-9**
  - internal reference voltage, **23-3**
  - multislope, **23-8**
  - noise rejection by, **23-10**
  - nonintegrating, **23-9**
  - resistance measurement with, **23-6**
  - resolution of, **7-22**
  - sensitivity of, **7-18, 7-22**
  - successive approximation, **23-9**
- Digital multimeters (*Cont.*):
  - temperature coefficient and, theory of operation of, **7-20**
- Digital storage oscilloscopes, **24-30 to 24-33**
- Digital-to-analog (D/A) conversion, **15-20 to 15-23**
- Diode logic, **14-7**
- Diode stabilization, **12-9**
- Diodes:
  - avalanche-type, **8-29**
  - clamping, **15-5**
  - double-anode regulator, **8-9**
  - forward-biased, **8-5**
  - integrated circuits and, **9-11**
  - junction (*see* Junction diodes)
  - laser, **21-28 to 21-32**
  - light-emitting (LED), **7-19, 8-25, 21-23 to 21-28**
  - microwave power, **8-29**
  - photo-, **8-27**
    - avalanche, **21-32**
  - pn, **21-32**
  - radio-frequency (r-f), **8-7**
  - rectifier, **8-7**
  - reference, **8-9**
  - regulator, **8-9**
  - reverse-biased, **8-5, 21-32**
  - Schottky barrier, **8-27**
  - switching, **8-7, 14-4, 14-17**
  - transferred-electron bulk, **8-29**
  - tunnel, **8-26**
  - zener, **8-8**
- DIP (dual-in-line package) resistors, **1-13**
- Direct addressing, **18-11, 18-12**
- Direct-current (dc) coil inductance, **3-23**
- Direct-current (dc) input amplification, **23-2**
- Direct-current (dc) input attenuation, **23-2**
- Direct-current (dc) leakage, **2-5**
- Direct-current (dc) level shifter, **13-13**
- Direct-current (dc) (average) value of periodic waveforms, **6-19**
- Directivity, **20-24, 20-34, 20-35**
- Discretionary wiring, **9-16**
- Display modes, logic analyzer, **19-8, 19-17 to 19-20**
  - binary, **19-9, 19-17, 19-27**
  - compare, **19-18, 19-27**
  - hexadecimal, **19-9, 19-17, 19-27**
  - inverted-logic, **19-27**
  - map, **19-9, 19-19, 19-27**
  - terminology for describing, **19-27**
  - time, **19-8, 19-18**
- Display window, **19-18, 19-27**
- Displays, digital, **7-19**

- Dissipation:
  - collector, 12-23
  - power, 1-2
- Dissipation factor, 2-6, 7-31
- Distortion, 12-28
  - crossover, 12-25
  - harmonic, 12-4, 12-24
  - intermodulation, 12-4
  - second-harmonic, 12-24
- Distributed capacitance, 3-11 to 3-13
  - measuring, 3-15
- Dopant, 8-4, 9-6
- Doping, 8-4
- Dot notation, 5-9
- Double-anode regulator diodes, 8-9
- Double-tuned circuits, 10-4, 10-18 to 10-20
- Down counter, 15-14
- Drain characteristics, 8-15
- Drain dissipation, 12-23
- Drift, 1-5
- Driven antenna arrays, 20-36
- Dry-cell batteries (*see* Primary batteries)
- Dual clocking, 19-27
- Dual-gang capacitors, 2-11
- Dual-in-line package (DIP) resistors, 1-13
- Dual-slope integration, 23-7
- Dual-slope multimeters, 23-7
- Dual-time-base oscilloscopes, 24-22
- Dummy load, 7-8
- Dynamic response, 14-1
  
- Eccles-Jordan circuits, 14-14
- ECL (emitter-coupled logic) circuits, 15-6
- Eddy-current loss in transformers, 5-15
- Eddy currents, 3-10, 3-11, 4-19
- Edge triggering, 19-26
- Effective area, antenna, 20-35
- Effective (rms) value of periodic waveforms, 6-19
- 8080 microprocessors, 18-24
- 8085 instruction set, 18-13 to 18-19
  - arithmetic group, 18-16
  - branch control group, 18-18
  - data transfer group, 18-13
  - input-output operations, 18-18
  - logic group, 18-17
  - machine control group, 18-18
  - stack operations, 18-18
- 8085 microprocessors, 18-6 to 18-12
  - address buffer, 18-8
  - address latch, 18-8
  - addressing, 18-11, 18-12
  - arithmetic logic unit (ALU), 18-8
  - 8085 microprocessors (*Cont.*):
    - instruction decoder, 18-8
    - instruction register, 18-8
    - instruction set for (*see* 8085 instruction set)
    - interrupt control, 18-8
    - multiplexing, 18-10
    - program counter, 18-7
    - registers, 18-6
    - serial input-output (I/O) control, 18-9
- Electrodes, 2-1
- Electrolytic capacitors, 2-9
- Electromagnetism, 4-5 to 4-7, 4-17
  - left-hand rule and, 4-6
  - right-hand rule and, 4-6
- Electronic filters, 16-18
- Emitter-coupled logic (ECL) circuits, 15-6
- End resistance, 1-6
- Energy in electric circuits, 6-5
- Energy bands, 8-3
- Energy-storage capacitors, 2-9
- Enhancement-type MOSFET, 8-17, 12-11
- Epitaxial layer, 9-3
- Equalizing resistors, 2-11
- Equivalent resistance, 5-18
- Equivalent series resistance, 2-6
- Etching, 9-7
- Evanescence field, 21-9
- EXCLUSIVE OR gate, 14-11
- Extended addressing, 18-23
- External triggering, 19-15, 19-26
  
- Fall time, 14-19
- Fan in, 15-1
- Fan out, 15-3
- Far field antenna pattern, 20-33
- Fast-transfer cathode-ray tube (CRT)
  - oscilloscopes, 24-29
- Feedback, 13-2
  - current, 12-28
  - voltage, 12-8, 12-28
- Feedback amplifier, 12-26 to 12-28
- Ferrite core coils, 3-1
- Ferromagnetism, 4-1
- Ferrous core coils, 3-1 to 3-3
  - ferrite, 3-1
  - iron-powder, 3-1
  - laminated, 3-1
  - permeability of, 3-3
  - toroidal, 3-3
  - tuning, 3-3
- FET (field-effect transistor), 8-13 to 8-18, 9-11
  - amplifying, 12-15 to 12-17

- FET (field-effect transistor) (*Cont.*):  
 integrated circuit, 9-11  
 junction (JFET), 8-13, 12-9  
 MOS (*see* MOSFET)  
 terminology for describing, 8-18
- Fiber-optic communications, 21-1 to 21-38  
 advantages of, 21-2 to 21-4  
 fibers for (*see* Optical fibers)  
 optical receivers in (*see* Optical receivers)  
 transmitters in (*see* Optical transmitters)
- Field-effect transistor (*see* FET)
- Filter chokes, 11-3
- Filters, 11-1 to 11-29  
 active (*see* Active filters)  
 amplifying, 22-3  
 applications of, 11-2  
 band-reject, 11-14, 11-29  
 bandpass, 11-12 to 11-14, 11-29, 22-17  
 bandstop, 11-14, 22-20  
 Bessel, 22-5  
 broadband, 22-16  
 Butterworth, 11-22, 11-26 to 11-28, 22-5  
 cascaded, 22-13, 22-15 to 22-16  
 ceramic, 11-23  
 Chebyshev, 11-22, 22-5  
 classification of, 11-1  
 components, 11-6  
 constant- $k$  (*see* Constant- $k$  filters)  
 crystal, 11-22  
 first-order, 22-2  
 high-pass (*see* High-pass filters)  
 image-parameter design method, 11-6  
 impedance and, 11-7 to 11-9  
 low-pass (*see* Low-pass filters)  
 $m$ -derived (*see*  $M$ -derived filters)  
 magnetostrictive, 11-25  
 mechanical, 11-25  
 narrowband, 22-17  
 network design method, 11-7  
 rectifier, 16-3  
 resistive-reactive, 11-6  
 second-order (*see* Second-order filters)  
 terminology for describing, 11-4 to 11-6  
 third-order, 22-12 to 22-16  
 totally reactive, 11-6
- First-order filters, 22-2
- Fixed capacitors, 2-7 to 2-11
- Fixed resistors, 1-8 to 1-11
- Fixed-tuned circuits, 10-3
- Flag flipflop, 18-7
- Flip chips, 8-30, 9-19
- Flipflops, 15-8 to 15-13  
 clearing, 15-12  
 clocked R-S (reset-set), 15-9
- Flipflops (*Cont.*):  
 delay, 15-12  
 flag, 18-7  
 $J$ - $K$ , 15-10  
 master-slave, 15-13  
 presetting, 15-12  
 reset-set (R-S), 15-8  
 $T$ , 15-11  
 (*See also* Counters)
- Focusing antennas, 20-40
- Formattable memory, 19-27
- Forward-biased diodes, 8-5
- Forward transconductance, 12-18
- Four-point probes, 9-12
- Frequency:  
 break, 13-7  
 center, 11-5  
 crossover, 13-7  
 cutoff, 2-16, 11-7  
 fundamental, 12-4  
 infinite attenuation, 11-16  
 Nyquist, 24-31  
 ripple, 5-35
- Frequency effects, 1-4
- Frequency response, 12-17 to 12-21, 13-6  
 cascaded amplifiers, 12-22  
 figures of merit for, 12-19  
 high-frequency model, 12-17, 12-20  
 hybrid- $\pi$  model, 12-17 to 12-19  
 low-, 12-21
- Frequency stability, 12-30
- Fuel cells, 17-31 to 17-33
- Full-wave rectifiers, 16-4
- Full-wave voltage doublers, 16-10
- Fundamental frequency, 12-4
- Gain:  
 antenna, 20-34  
 closed-loop, 12-27, 13-7  
 current, 12-3  
 open-loop, 12-27, 13-6  
 power, 12-3  
 voltage, 12-3
- Gain stability, 12-27
- Ganged variable resistors, 1-13
- Gate dissipation, 15-3
- Gauss (unit), 4-8, 4-11
- Gilbert (unit), 4-9, 4-11
- Glass capacitors, 2-8
- Glitch capturing mode, 19-21, 19-27
- Glitch triggering, 19-26
- Glitches, 19-3
- Graded-index optical fibers, 21-7



- Grain-oriented laminations, transformer, 5-40
- Grand-scale integration (GSI), 15-1
- Ground point, 7-4
- Half-adder, 14-10
- Half-wave rectifiers, 16-2
- Half-wave voltage doublers, 16-9
- Harmonic distortion, 12-4, 12-24
- Hartley oscillators, 12-30
- Heat sinks, 12-25
- Hexadecimal display mode, 19-9, 19-17, 19-27
- High-frequency model, 12-17, 12-20
- High-megohm resistors, 1-13
- High-pass filters, 11-11, 11-18, 11-28, 22-2 to 22-16, 22-24
  - first-order, 22-2
  - second-order (*see* Second-order filters)
  - third-order, 22-12 to 22-16
- High-speed TTL, 15-5
- High-voltage resistors, 1-13
- Hold time, 19-11, 19-27
- Horn antennas, 20-40
- Hybrid model, 12-11
- Hybrid-pi model, 12-17 to 12-19
- Hybrids, 20-24
- Hysteresis, 4-19
- Hysteresis loss, 4-19, 5-15
- IC (*see* Integrated circuits)
- Ideal element, 6-1
- Image impedance, 11-9
- Image-parameter design method, 11-6
- Immediate addressing, 18-11, 18-12
- Impedance, 10-1
  - in capacitors, 2-6
  - characteristic, 11-9, 20-3
  - circuit analysis and, 6-22
  - equivalent, 6-26
  - filter, 11-7 to 11-9
  - image, 11-9
  - surge, 11-9
- Implied addressing, 18-24
- Impurity concentrations, 9-6
- Indexed addressing, 18-11, 18-24
- Indirect addressing, 18-11, 18-12
- Inductance:
  - critical, 16-20
  - direct current and coil, 3-23, 3-24
  - mutual (*see* Mutual inductance)
  - standard, 7-29
- Inductive reactance, 10-1, 10-6
- Inductors, 6-2, 9-18
  - integrated circuit, 9-13
  - reactance of, 10-1, 10-6
- Infinite attenuation frequency, 11-16
- Infinite gain multiple feedback (IGMF) circuits, 22-6
- INHIBIT gate, 14-11
- Input offset current, 13-8
- Input offset voltage, 13-8
- Input-output (I/O) instructions, 18-18
- Input-output (I/O) ports, 18-22
- Input resistance, 12-3, 12-28
- Insertion loss, 11-4
- Instruction decoder, 18-8
- Instruction register, 18-8
- Instrument transformers, 5-26 to 5-28
- Instrumentation, 7-1 to 7-22
  - accuracy of, 7-3, 7-22
  - ammeters, 7-3
  - digital multimeters (*see* Digital multimeters)
  - meter circuits (*see* Meter circuits)
  - meter movements (*see* Meter movements)
  - meter resistance, 7-13
  - ohmmeters, 7-5
  - parallax and, 7-1
  - sensitivity of, 7-18, 7-22
  - voltmeters, 7-4, 7-16 to 7-18
- Insulation resistance, 2-5
- Integrated circuits (IC), 9-1
  - beam-lead isolation, 9-9
  - bipolar junction transistor (*see* Bipolar junction transistors)
  - capacitors and, 9-12
  - dielectric isolation, 9-9
  - digital (*see* Digital integrated circuits)
  - dimensional units for, 9-1
  - diode, 9-11
  - FET, 9-11
    - (*See also* FET)
  - four-point probe, 9-12
  - inductors and, 9-13
  - junction isolation, 9-9
  - lateral pnp transistor, 9-10
  - manufacture of (*see* Circuit integration)
  - parasitics, 9-10
  - resistor, 9-11
- Integration:
  - dual-slope, 23-7
  - grand-scale (GSI), 15-1
  - large-scale (LSI), 9-16, 15-1
  - medium-scale (MSI), 15-1
  - multislope, 23-8



- Integration (*Cont.*):
  - small-scale (SSI), 15-1
- Interbase resistance, 16-12
- Interconnection mask, 9-9, 9-14
- Intermodulation distortion, 12-4
- Internal triggering, 19-13
- Interrupt control, 18-8
- Intrinsic standoff ratio, 8-22
- Inverted-logic display mode, 19-27
- INVERTER, 14-8, 15-7
- Inverting amplifiers, 13-2
- I/O (*see* Input-output *entries*)
- Iron-powder core coils, 3-1
- Isolation, 20-24
  - beam-lead, 9-9
  - dielectric, 9-9
  - junction, 9-9
- Isolation mask, 9-14
- Isolation transformers, 5-21
- Isolator, 20-25
  - opto-, 8-28
- Isotropic antennas, 20-33
- Isotropic radiators, 20-33
  
- J-K flipflop, 15-10
- Jeweled-type meter movements, 7-9
- JFET (junction FET), 8-13, 12-9
- JFET amplifiers, 12-9
- Joule (unit), 6-5
- Joule effect, 11-25
- Junction capacitors, 9-12
- Junction diodes, 8-5 to 8-8, 9-11
  - characteristics of, 8-5
  - radio-frequency (r-f), 8-7
  - rectifier, 8-7
  - switching, 8-7
  - terminology for describing, 8-7
- Junction FET (JFET), 8-13, 12-9
- Junction isolation, 9-9
- Junctions:
  - base-emitter, 8-9
  - collector-base, 8-9
  - isolation, 9-9
  - pn, 8-5
  
- Kirchhoff's laws, 6-5
- Kodak Photo Resist (KPR), 9-4
  
- Ladder circuits, 11-4
- Ladder-type converters, 15-21
- Lambertian distribution, 21-23
- Laminated core coils, 3-1
- Large-scale integration (LSI), 9-16, 15-1
- Laser diodes, 21-28 to 21-32
- Latch, 15-9
- Lateral pnp transistors, 9-10
- Lattice circuits, 11-4
- Lead-acid batteries, 17-17 to 17-19, 17-24
- Leadless inverted device (LID), 8-30
- LeClanche, Georges, 17-1
- LeClanche batteries, 17-5 to 17-11
  - (*See also* Primary batteries)
- LED (light-emitting diodes), 7-19, 8-25, 21-23 to 21-28
- Left-hand rule, 4-6
- LID (leadless inverted device), 8-30
- Light-activated thyristors, 8-20
- Light-emitting diodes (LED), 7-19, 8-25, 21-23 to 21-28
- Light-guiding optical fibers, 21-4 to 21-7
- Limiters, 20-28
- Linear circuits, 6-3
- Linear element, 6-3
- Linear resistors, 1-1
- Linear taper, 1-6
- Lissajou patterns, 24-19
- Load, power supply, 16-12
- Load conditions, transformer, 5-11
- Load current, 5-11
- Load resistance, 16-12
- Logarithmic amplifiers, 13-5
- Logic:
  - negative, 14-7
  - positive, 14-7
- Logic analyzers, 19-6 to 19-29
  - data acquisition function of (*see* Data acquisition, logic analyzer)
  - data transport in, 19-21 to 19-26
  - display function of (*see* Display modes, logic analyzer)
  - memory of, 19-6, 19-27, 19-28
  - required features of, 19-6
  - terminology for describing, 19-26 to 19-29
  - trigger function of (*see* Trigger function, logic analyzer)
- Logic circuits, 14-6 to 14-14
  - AND gate, 14-7
  - EXCLUSIVE OR gate, 14-11
  - half-adder, 14-10
  - INHIBIT gate, 14-11
  - INVERTER, 14-8, 15-7
  - NAND gate, 14-12, 15-8
  - NOR gate, 14-13, 15-7
  - OR gate, 14-6
- Logic instructions, 18-17

- Logic probes, 19-30
- Logic pulsers, 19-30
- Loop gain, 12-27
- Low-pass filters, 11-9 to 11-11, 11-17, 22-2 to 22-16
  - first-order, 22-2
  - second-order (*see* Second-order filters)
  - third-order, 22-12 to 22-16
- Low-power TTL circuits, 15-4
  
- m*-derived filters, 11-7, 11-15 to 11-21
  - design of, 11-19 to 11-21
  - high-pass, 11-18
  - low-pass, 11-17
- Machine control instructions, 18-18
- Magnesium dry batteries, 17-15, 17-28
- Magnetic circuits, 4-1 to 4-21
  - coercivity and, 4-18
  - compared to electric circuits, 4-9
  - Coulomb's law and, 4-4
  - electromagnetism and (*see* Electromagnetism)
  - flux density and, 4-8
  - left-hand rule and, 4-6
  - magnetic flux and, 4-7
  - magnetizing force and, 4-10
  - magnetomotive force and, 4-8
  - materials of, 4-1
  - Ohm's laws for, 4-11
  - relative permeability and, 4-11
  - reluctance in (*see* Reluctance)
  - right-hand rule and, 4-6
  - units of measure of (*see* Magnetic units) (*See also* Magnetic poles)
- Magnetic compass, 4-3
- Magnetic declination, 4-5
- Magnetic field, terrestrial, 4-5
- Magnetic flux, 4-7
- Magnetic flux density, 4-8
- Magnetic poles, 4-3 to 4-6
  - computation of, 4-4
  - Coulomb's law and, 4-4
  - terrestrial, 4-5
  - unit, 4-4, 4-7
- Magnetic units, 4-7 to 4-11
  - and electric units, 4-9
  - for flux density, 4-8
  - for magnetic flux, 4-7
  - for magnetizing force, 4-10
  - for magnetomotive force, 4-8
  - for relative permeability, 4-11
  - for reluctance, 4-9
  - for unit poles, 4-7
- Magnetite, 4-1
- Magnetizing current, 5-9
- Magnetizing force, 4-10
- Magnetomotive force, 4-8
- Magnetostrictive filters, 11-25
- Magnets, 4-1
- Majority carrier device, 8-27
- Manual triggering, 19-13, 19-26
- Map display mode, 19-9, 19-19, 19-27
- Masks, integrated circuit:
  - base, 9-8
  - collector, 9-6
  - contact, 9-8, 9-14
  - interconnection, 9-9, 9-14
  - isolation, 9-14
  - n+, 9-8, 9-14
  - p, 9-14
- Master-slave flipflop, 15-13
- Maxwell (unit), 4-7, 4-11
- Mechanical filters, 11-25
- Medium-scale integration (MSI), 15-1
- Memories (*see* Semiconductor memories)
- Memory, logic analyzer, 19-6, 19-27, 19-28
  - formattable, 19-27
- Memory depth, 19-28
- Memory width, 19-28
- Mercury dry batteries, 17-13 to 17-15
- Mesh count, 9-19
- Metal-film resistors, 1-8
- Metal-glaze resistors, 1-8
- Metal-oxide semiconductor field-effect transistor (*see* MOSFET)
- Metallization, 9-8
- Metallized dielectric capacitors, 2-9
- Meter circuits, 7-11 to 7-16
  - current in, 7-13
  - current-reading meters and, 7-13
  - meter multiplier, 7-14 to 7-16
  - meter shunt, 7-11 to 7-13
  - resistance in, 7-13
- Meter current, 7-13
- Meter movements:
  - jeweled-type, 7-9
  - moving-coil, 7-9 to 7-11
  - taut-band, 7-9
- Meter multipliers, 7-14 to 7-16
- Meter resistance, 7-13
- Meter shunts, 7-11 to 7-13
- Mica capacitors, 2-7 to 2-8
- Microcomputers, 18-5
  - (*See also* Microprocessors)
- Microprocessors, 18-1 to 18-26
  - accumulator addressing, 18-24
  - digital words, 18-1 to 18-3

Microprocessors (*Cont.*):

- 8080, 18-24
- 8085 (*see* 8085 microprocessors)
- extended addressing, 18-23
- implied addressing, 18-24
- indexed addressing, 18-11, 18-24
- microcomputers and, 18-5
- programming, 18-19 to 18-21
- relative addressing, 18-24
- 6502, 18-25
- 6800, 18-25
- support devices for, 18-22
- Z80, 18-24

## Microstrips, 20-5

## Microwave amplifiers, 20-29

## Microwave circuit components, 20-21 to 20-33

- active linear nonreciprocal, 20-29
- active linear reciprocal, 20-29
- passive linear nonreciprocal, 20-25
- passive linear reciprocal, 20-22 to 20-25
- passive nonlinear reciprocal, 20-26 to 20-28

## Microwave power diodes, 8-29

## Microwave region, 20-1

## Microwave test equipment:

- insertion loss test set, 20-45
- network analyzer, 20-47
- power meter, 20-41

## Microwaves, 20-1 to 20-42

## antennas for, 20-38 to 20-40

(*See also* Antennas)

circuit components for (*see* Microwave circuit components)

- decibels and, 20-2
- detection of, 20-26
- power and, 20-2
- region of, 20-1
- test equipment for (*see* Microwave test equipment)
- transmission lines for (*see* Transmission lines, microwave)

## Miller capacitance, 12-18

## Miller effect, 12-18

## Mixers, 20-26

## Model, 6-2

## Modulo counter, 15-14

## Monostable multivibrators, 14-16

## MOSFET (metal-oxide semiconductor FET),

8-15 to 8-17

- amplifiers (*see* MOSFET amplifiers)
- depletion-type, 8-15, 12-9 to 12-11
- enhancement-type, 8-17, 12-11

## MOSFET amplifiers:

- analyzing, 12-15 to 12-17
- biasing, 12-9 to 12-11
- depletion-type, 12-9 to 12-11
- enhancement-type, 12-11

## Moving-coil meter movements, 7-9 to 7-11

Multimeters, digital (*see* Digital multimeters)

## Multimode guide, 21-7

## Multiple-tuned circuits, 10-4, 10-18 to 10-20

## Multiple-turn variable resistors, 1-13

## Multiplexing, 18-10

## Multislope integration, 23-8

## Multislope multimeters, 23-8

## Multivibrators (MV), 14-14 to 14-17

- astable, 14-17
- bistable, 14-15
- monostable, 14-16

## Mutual conductance, 6-8

## Mutual inductance, 3-5 to 3-9, 5-1 to 5-4, 10-20

- air-core coil, 3-7
- measuring, 3-9
- parallel coil, 3-6
- series coil, 3-6
- transformer, 5-1 to 5-4

## Mutual resistance, 6-7

MV (*see* Multivibrators)

## n-type semiconductors, 8-4

## n+ buried layer, 9-10

## n+ diffusion, 9-8, 9-14

## n+ mask, 9-8, 9-14

## NAND gate, 14-12, 15-8

## Narrowband filters, 22-17

## Negative and positive voltage output, 16-12

## Negative-going edge, 15-14

## Negative logic, 14-7

## Negative reactance, 10-1

## Negative resistance, 8-22

## Network method of design, 11-7

## Networks, topology of, 6-8

## Nibble, 18-2

## Nickel-cadmium batteries, 17-19 to 17-22, 17-26

## No-load conditions, transformer, 5-8 to 5-11

## No-load current, 5-8

## Noise, 1-4

## Noise equivalent power, 21-36 to 21-38

## Noise margin, 15-3

## Nomograms, 3-24

## Nonferrous core coils, 3-3

- Noninverting amplifiers, 13-3
- Nonlinear circuits, 6-3
- Nonlinear element, 6-3
- Nonlinear resistors, 1-1, 1-16
- Nonlinear taper, 1-6
- Nonpolarized electrolytic capacitors, 2-9
- Nonsaturated switch, 14-5
- NOR gate, 14-13, 15-7
- Nyquist frequency, 24-31
  
- Oersted (unit), 4-11
- Offset, 13-8
- Offset correction, 13-8
- Ohm's law, 1-2
- One-to-one transformers, 5-1
- Open-circuit test, 5-16
- Open circuits, 6-3
- Open-loop gain, 12-27, 13-6
- Operational amplifiers (op amp), 13-1 to 13-18
  - compensated, 13-6
  - current-to-voltage converters and, 13-6
  - difference, 13-3
  - differential, 13-10
  - direct-current level shifter in, 13-13
  - feedback in, 13-2
  - frequency response of, 13-6
  - logarithmic, 13-5
  - noninverting, 13-3
  - offset in, 13-8
  - output stage of, 13-13
  - slewing rate of, 13-9
  - summing, 13-4
  - third-order filter, 22-12
  - voltage follower in, 13-4
- Optical fiber communications (*see* Fiber-optic communications)
- Optical fibers, 21-4 to 21-21
  - attenuation in, 21-7 to 21-12
  - bandwidth of, 21-12 to 21-14
  - cabled, 21-14 to 21-17
  - connecting mismatch loss in, 21-20
  - connectors for, 21-18
  - coupling, 21-19
  - graded-index, 21-7
  - light-guiding, 21-4 to 21-7
  - splicing of, 21-17
  - step-index, 21-5
- Optical receivers, 21-32 to 21-38
  - noise equivalent power of, 21-36 to 21-38
  - responsivity of, 21-34 to 21-36
  - sensitivity of, 21-38
- Optical transmitters, 21-21 to 21-32
  - laser diode, 21-28 to 21-32
  - light-emitting diode, 21-23 to 21-28
- Opto-isolator, 8-28
- OR gate, 14-6
- Organic depolarized batteries, 17-28
- Oscillators, sinusoidal, 12-28 to 12-31
  - Colpitts, 12-30
  - crystal, 12-31
  - frequency stability of, 12-30
  - Hartley, 12-30
  - phase-shift, 12-29
  - tuned-circuit, 12-30
- Oscilloscope measurements, 24-11 to 24-28
  - bandwidth and, 24-25
  - basic waveforms of, 24-13
  - circuit loading and, 24-27
  - delayed-sweep, 24-21 to 24-24
  - differential, 24-20
  - direct, 24-14
  - frequency, 24-16
  - indirect, 24-15
  - period and, 24-16
  - phase, 24-18
  - probes, 24-12
  - pulse, 24-17
  - rectangular-wave, 24-17
  - risetime and, 24-24
  - safety in, 24-11
  - setup, 24-11
  - XY, 24-19
  - Z axis, 24-21
  - (*See also* Oscilloscopes)
- Oscilloscopes, 24-1 to 24-11, 24-28 to 24-35
  - bandwidth and, 24-15, 24-25, 24-27
  - cathode-ray tube storage, 24-28 to 24-30
  - coupling in, 24-5, 24-11
  - digital storage, 24-30 to 24-33
  - display system of, 24-2 to 24-4
  - dual-time-base, 24-22
  - horizontal controls of, 24-6
  - selection of, 24-33
  - single-time-base, 24-24
  - trigger holdoff, 24-8
  - trigger operating modes, 24-10
  - trigger point, 24-8
  - trigger sources, 24-9
  - vertical controls of, 24-4 to 24-6
  - (*See also* Oscilloscope measurements)
- Output resistance, 12-3, 12-28
- Oxide layer, 9-3

- p-diffusion, 9-7, 9-14
- p-mask, 9-14
- p-type semiconductors, 8-5
- p-type substrate, 9-3
- Paper capacitors, 2-8
- Parallax, 7-1
- Parallel capacitors, 2-10
- Parallel circuits, 6-3
- Parallel-in, parallel-out registers, 15-15
- Parallel-in, serial-out registers, 15-15
- Parallel resistors, 1-2
- Parallel resonance, 10-11
- Parallel transformers, 5-14
- Parallel-tuned circuits, 10-10 to 10-15
- Parallel-wire transmission lines, 20-5
- Paramagnetism, 4-3
- Parameter field, 19-27
- Parasitic antenna arrays, 20-36
- Parasitics, 1-5, 9-10, 11-6
- Pass counter, 19-26
- Passband, 11-4
- Passive element, 6-1
- Pattern multiplication, 20-37
- Peak current, 8-23
- Peak voltage, 8-23
- Periodic waveforms, 6-15 to 6-19
  - average (dc) value of, 6-19
  - effective (rms) value of, 6-19
- Permeability of ferrous core coils, 3-3
- Permeance, 4-9, 4-13
- Permittivity, 2-1
- Phase-inverting input transformers, 12-25
- Phase-shift oscillators, 12-29
- Phase shifters, 20-24
- Phase splitters, 13-13, 15-3
- Photodiodes, 8-27
  - avalanche, 21-32
- Photoresist, 9-4
- Phototransistors, 8-28, 11-32
- Piezoelectric effect, 12-31
- Plante, Gaston, 17-1
- Plastic film capacitors, 2-8
- pn diodes, 21-32
- pn junctions, 8-5
- Polarization, 20-34
  - of dry cells, 17-5
- Polarized antennas, 20-34
- Polarized electrolytic capacitors, 2-9
- Poles, magnetic (*see* Magnetic poles)
- Polycarbonate capacitors, 2-8
- Polyester capacitors, 2-8
- Polystyrene capacitors, 2-15
- Porcelain chip capacitors, 2-15
- Positive and negative output voltage, 16-12
- Positive logic, 14-7
- Positive reactance, 10-1
- Post-triggering, 19-26
- Potential transformers, 5-27
- Potentiometers (*see* Variable resistors)
- Power, in electric circuits, 6-5
- Power amplifiers, 12-22 to 12-26
  - class A, 12-23
  - collector dissipation in, 12-23
  - crossover distortion of, 12-25
  - heat sinks and, 12-25
  - push-pull, 12-24
  - second-harmonic distortion of, 12-24
- Power dissipation, resistor, 1-2, 1-3
- Power factor (PF), 2-5
- Power gain, 12-3
- Power rating, 1-3
- Power-style transistors, 1-10
- Power supplies, 16-1 to 16-18
  - rectifiers and (*see* Rectifiers)
  - regulated, 16-13 to 16-15
  - (*See also* Regulators)
- Power transistors, 20-19
- Precision-style resistors, 1-10
- Pretriggering, 19-26
- Primary batteries, 17-2, 17-5 to 17-16
  - alkaline, 17-11 to 17-13
  - American National Standards Institute (ANSI) standards for, 17-10
  - LeClanche, 17-5 to 17-11
  - magnesium dry, 17-15, 17-28
  - mercury dry, 17-13 to 17-15
  - organic depolarized, 17-28
  - performance of, 17-6 to 17-10
  - polarization of, 17-5
  - shelf life of (*see* Shelf life)
- Primary winding, 5-1
- Probe compensation, 24-12
- Probes, 9-9, 24-12
  - four-point, 9-12
  - logic, 19-30
- Program counter, 18-7
- Program status word, 18-7
- Programmable direct memory access (DMA)
  - controller, 18-22
- Programmable interrupt controller, 18-22
- Programmable keyboard interface, 18-22
- Programmable peripheral interface (PPI), 18-22
- Programmable unijunction transistors (PUT), 8-24
- Programmable video display controller, 18-22
- Programming, 18-19 to 18-21



- Propagation delay, 15-3
- Pulsating direct-current (dc) voltage, 16-2
- Punch-through transistors, 9-10
- Push-pull power amplifiers, 12-24
  
- Quality factor:
  - active filter, 22-5
  - capacitor, 2-7
  - tuned circuit, 10-16 to 10-18
- Quiescent operating point (*Q* point), 12-5
  
- Race condition, 15-13
- Races, 19-4
- Radio-frequency (r-f) diodes, 8-7
- Random-access memory (RAM), 15-19
- Rated continuous working voltage (RCWV), 1-4
- RC circuits, 6-3
- Reactance:
  - capacitive, 2-5, 10-1, 10-6
  - equivalent, 6-23
  - inductive, 10-1, 10-6
  - negative, 10-1
  - positive, 10-1
- Read-only memory (ROM), 15-18
- Receivers, optical (*see* Optical receivers)
- Rechargeable batteries (*see* Secondary batteries)
- Reciprocal antennas, 20-33
- Rectifiers, 16-1 to 16-13
  - bridge, 16-7
  - diode, 8-7
  - filtering and, 16-3
  - full-wave, 16-4
  - full-wave voltage doublers and, 16-10
  - half-wave, 16-2
  - half-wave voltage doublers and, 16-9
  - load and, 16-12
  - load resistance and, 16-12
  - negative and positive output voltage, 16-12
  - tube, 16-6
  - voltage triplers and, 16-10
- Reference diodes, 8-9
- Reflection coefficient, 20-13
- Registers, 18-6, 18-8
  - circulating, 15-16
  - serial, 15-15
  - shift, 15-15
- Regulated power supplies, 16-13 to 16-15
  - (*See also* Regulators)
- Regulators, 16-13 to 16-18
  - circuitry of, 16-16 to 16-18
  - closed-loop, 16-17
  - diode, 8-9
  - series, 16-15
  - shunt, 16-15
  - switching of, 16-15
- Relative addressing, 18-24
- Relative dielectric constant, 2-3
- Relative permeability, 4-11
- Reluctance, 4-9
- Reserve batteries, 17-3, 17-29
- Reset-set (R-S) flipflop, 15-8
- Residual magnetism, 4-18
- Resistance, 1-1, 1-3
  - base-emitter, 12-17
  - base-spreading, 12-17
  - collector-emitter, 12-18
  - contact, 1-6
  - critical, 1-4
  - end, 1-6
  - input, 12-3, 12-28
  - insulation, 2-5
  - interbase, 16-12
  - load, 16-12
  - meter, 7-13
  - mutual, 6-7
  - negative, 8-22
  - output, 12-3, 12-28
  - self-, 6-7
  - series, 2-16
  - equivalent, 2-6
  - sheet, 9-12
  - temperature coefficient of, 1-3
- Resistive-reactive filters, 11-6
- Resistors, 1-1 to 1-19, 6-1, 9-11, 9-17
  - carbon composition, 1-8
  - carbon film, 1-9
  - chip, 1-16
  - color coding of, 1-11
  - drift in, 1-5
  - dual-in-line package (DIP), 1-13
  - equalizing, 2-11
  - fixed, 1-8 to 1-11
  - frequency effects in, 1-4
  - high-megohm, 1-13
  - high-voltage, 1-13
  - integrated circuit, 9-11
  - linear, 1-1
  - metal-film, 1-8
  - noise in, 1-4
  - nonlinear, 1-1, 1-16
  - parallel, 1-12
  - power dissipation in, 1-2, 1-3



Resistors (*Cont.*):

- power rating, 1-3
- rated continuous working voltage of, 1-4
- series, 1-11, 6-9
- termination, 20-22
- tolerance in, 1-3
- variable (*see* Variable resistors)
- wire-wound, 1-10
- (*See also* Resistance)

## Resolution, 1-6

- instrumentational, 7-22

## Resonant frequency, 10-1, 10-7

## Responsivity, optical receiver, 21-34 to 21-36

## Reverse-biased diodes, 8-5, 21-32

## Reverse-recovery time, 14-18

## Rheostat, 1-16

## Right-hand rule, 4-6

## Ringing, 15-5, 19-4

## Ripple, 16-22

## Ripple counter, 15-13

## Ripple current, 2-7

## Ripple frequency, 5-35

## Risetime, 14-18, 24-24

## RL circuits, 6-3

Roll-off, constant- $k$  filter, 11-15

## ROM (read-only memory), 15-18

## Root-mean-square (rms) converter, 23-4

## Root-mean-square (rms) (effective) value of waveforms, 6-19

## Rotor, 2-11

## Rubyliith, 9-5

## S parameters, 20-19

## Sample interval, 19-27

## Sampling:

- asynchronous, 19-20, 19-27
- synchronous, 19-20, 19-27

## Saturated switch, 14-5

## Saturation, magnetic, 4-17

## Saturation region, 14-5

## Saturation voltage, 14-5

## Schottky barrier diodes, 8-27

## Schottky-clamped TTL, 15-5

## Schottky transistors, 15-5

## Search mode, 19-28

## Second-harmonic distortion, 12-24

## Second-order filters, 22-5 to 22-12

- design data for, 22-9, 22-10
- effect of quality factor, 22-5
- single-feedback, 22-6

## Secondary batteries, 17-3, 17-17 to 17-27 charging, 17-24, 17-26

Secondary batteries (*Cont.*):

- lead-acid, 17-17 to 17-19, 17-24
- maintenance of, 17-24 to 17-27
- nickel-cadmium, 17-19 to 17-22, 17-26
- sealed unit, 17-25
- silver-zinc, 17-22 to 17-24, 17-26

## Secondary winding, 5-1

## Selectivity:

- filter, 11-5
- tuned-circuit, 10-4, 10-15

## Self-conductance, 6-8

## Self-heating voltage, 1-18

## Self-resistance, 6-7

## Self-resonant frequency, 3-14

## Semiconductor memories, 15-17 to 15-19

- access time, 15-18
- cycle time, 15-18
- random-access (RAM), 15-19
- read-only (ROM), 15-18

## Semiconductors, 8-1 to 8-9, 8-25 to 8-36

- beam-lead, 8-30
- charge-coupled device (CCD), 8-28
- doping, 8-4
- double-anode regulator diode, 8-9
- energy bands and, 8-3
- flip chip, 8-30, 9-19
- junction diode (*see* Junction diodes)
- leadless inverted device (LID), 8-30
- light-emitting diode (LED), 7-19, 8-25, 21-23 to 21-28

memories in (*see* Semiconductor memories)

- microwave power diode, 8-29
- n-type silicon, 8-4
- opto-isolator, 8-28
- p-type silicon, 8-5
- photodiode, 8-27
- phototransistor, 8-28, 21-32
- reference diode, 8-9
- regulator diode, 8-9
- Schottky barrier diode, 8-27
- theory of, 8-1 to 8-5
- tunnel diode, 8-26
- zener diode, 8-8
- (*See also* Transistors)

## Sensitivity:

- instrumentational, 7-18, 7-22
- optical receiver, 21-38

## Sequential triggering, 19-13, 19-26

## Serial-in, parallel-out registers, 15-15

## Serial input-output (I/O) control, 18-9

## Serial registers, 15-15

## Series capacitors, 2-10

## Series circuits, 6-3

- Series regulators, 16-15
- Series resistance, 2-16
  - equivalent, 2-6
- Series resistors, 1-11
- Series-tuned circuits, 10-5 to 10-10
  - applications of, 10-5
  - fundamentals of, 10-6 to 10-8
  - voltage of, at resonance, 10-8 to 10-10
- Series-tuned voltage, 10-8 to 10-10
- Serpentine resistor winding, 1-10
- Setting stability, 1-6
- Setup time, 19-11, 19-27
- Shape factor, 11-6
- Sheet resistance, 9-12
- Shelf life:
  - alkaline batteries, 17-12
  - dry batteries, 17-10
  - magnesium dry batteries, 17-15
  - mercury dry batteries, 17-13
- Shift registers, 15-15
- Short circuit, 6-3
- Short-circuit alpha cutoff frequency, 12-19
- Short-circuit beta cutoff frequency, 12-19
- Short-circuit current gain, 8-11
- Short-circuit gain-bandwidth product, 12-15
- Shunt, 6-9
- Shunt regulator, 16-15
- Signal-to-noise ratio (SNR), 21-34, 21-36 to 21-38
- Signature analysis, 19-30 to 19-33
  - design principles of, 19-32
  - ROM resident, 19-31
  - of unknown boards, 19-31
- Silver chloride-magnesium batteries, 17-29
- Silver-zinc batteries, 17-22 to 17-24, 17-26
- Silvered mica capacitors, 2-8
- Single-gang variable capacitors, 2-11
- Single-time-base oscilloscopes, 24-24
- Single-tuned circuits, 10-4
- Single-turn variable resistors, 1-13
- Sinusoidal oscillators (*see* Oscillators, sinusoidal)
- 6502 microprocessors, 18-25
- 6800 microprocessors, 18-25
- Skin effect:
  - in coils, 3-10
  - in transmission lines, 20-20
- Skirts, 11-15
- Slewing rate:
  - amplifier, 12-4
  - operational amplifier, 13-8
- Small-scale integration (SSI), 15-1
- Smith chart, 20-15 to 20-18
- Solid solubility, 9-6
- Source generator, 10-5
- Speed-power product, 15-3
- Spikes, 19-3
- Spiral antennas, 20-38
- Splices, optical fiber, 21-17
- Stability:
  - frequency, 12-30
  - gain, 12-27
- Stabilization:
  - bypass capacitors and, 12-9
  - common-emitter amplifier, 12-7 to 12-9
  - current-feedback, 12-8
  - decoupling capacitors and, 12-9
  - diode, 12-9
  - voltage-feedback, 12-8
- Stack instructions, 18-18
- Stack pointer, 18-7
- Stacking factor, 5-39
- Stagger-tuned circuits, 10-4
- Standard inductance, 7-29
- Standard resistance, 7-29
- Standing wave ratio (SWR) meter, 20-42
- Start function, 19-12
- State analyzers, 19-27
- Static emitter characteristics, 8-24
- Stator, 2-11
- Steady state, 6-3
- Step-and-repeat camera, 9-5
- Step-down transformers, 5-1
- Step-index optical fibers, 21-5
- Step-up transformers, 5-1
- Stop band, 11-6
- Storage time, 14-18
- Striplines, 20-5
- Subroutines, 18-21
- Successive approximation multimeters, 23-9
- Summing amplifiers, 13-4
- Super-beta transistors, 9-10
- Supply-voltage rejection ratio, 13-17
- Surge impedance, 11-9
- Switchers, 20-29
- Switching diodes, 8-7, 14-4, 14-17
- Switching levels, erroneous, 19-5
- Switching regulators, 16-15
- Synchronous counters, 15-14
- Synchronous sampling, 19-20, 19-27
- T* flipflop, 15-11
- Tantalum chip capacitors, 2-15
- Tantalum electrolytic capacitors, 2-9
- Taper, 1-5
- Taut-band meter movement, 7-9

- Temperature coefficient:  
  capacitance, 2-5  
  resistance, 1-3
- Temperature variation of capacitance, 2-16
- Termination resistors, 20-22
- Thermistors:  
  current-time characteristic of, 1-18  
  resistance temperature characteristic of, 1-16  
  static voltampere characteristic of, 1-17
- Thermocouple meters, 7-11
- Thermopile, 23-6
- Thin-film capacitors, 9-12
- Thin-film resistors, 1-18
- Third-order filters, 22-12 to 22-16
- Thyristors, 8-18 to 8-22  
  bidirectional triode, 8-21  
  characteristic curves of, 8-20  
  light-activated, 8-20  
  operation of, 8-19  
  terminology for describing, 8-21
- Time display mode, 19-8, 19-18
- Timing, microprocessor, 18-18, 18-19
- Timing analyzer speed, 19-27
- Timing analyzers, 19-27
- Timing errors, 19-4, 19-5
- Tin-oxide resistors, 1-8
- Tolerance:  
  capacitive, 2-4  
  resistive, 1-3
- Total diode capacitance, 2-16
- Totally reactive filters, 11-6
- Totem-pole amplifiers, 13-13, 15-4
- Transfer characteristics, 8-15
- Transfer function, 22-2
- Transfer switch, 20-29
- Transferred-electron bulk diodes, 8-29
- Transformer losses, 5-15 to 5-17  
  copper, 5-15  
  core, 5-16  
  eddy-current, 5-15  
  hysteresis, 5-15
- Transformers, 5-1 to 5-42  
  auto- (*see* Autotransformers)  
  basic, 5-1  
  current, 5-27  
  current relationships in, 5-5  
  efficiency of, 5-17  
  equivalent resistance in, 5-18  
  fundamental equation of, 5-6 to 5-8  
  instrument, 5-26 to 5-28  
  isolation, 5-21  
  load conditions in, 5-11  
  losses in (*see* Transformer losses)
- Transformers (*Cont.*):  
  mutual induction in, 5-1 to 5-4  
  no-load conditions in, 5-8 to 5-11  
  operation of, at abnormal frequency, 5-12 to 5-14  
  parallel, 5-14  
  potential, 5-27  
  series, 5-14  
  step-down, 5-1  
  step-up, 5-1  
  tuned filter chokes and, 5-36  
  voltage regulation of, 5-19 to 5-21
- Transimpedance amplifiers, 21-32
- Transistor switches, 14-4, 14-18
- Transistor-transistor logic circuits (*see* TTL circuits)
- Transistors, 8-9 to 8-24  
  bipolar junction (BJT) (*see* Bipolar junction transistors)  
  field-effect (*see* FET)  
  lateral pnp, 9-10  
  photo-, 8-28, 21-32  
  silicon-controlled rectifier (*see* Thyristors)  
  super-beta, 9-10  
  unijunction (UJT), 8-22 to 8-24  
  (*See also* Semiconductors)
- Transmission lines, microwave, 20-3 to 20-20  
  characteristic impedance of, 20-3, 20-8  
  coaxial cable, 20-4  
  types of, 20-6  
  dielectric, 20-13  
  microstrip, 20-5  
  parallel-wire, 20-5  
  power transfer in, 20-19  
  reflection coefficient and, 20-13  
  S parameters for, 20-19  
  skin effect in, 20-20  
  Smith chart for, 20-15 to 20-18  
  stripline, 20-5  
  voltage standing wave ratio (VSWR) in, 20-14  
  waveguides, 20-5
- Transmitters, optical (*see* Optical transmitters)
- Trigger function, logic analyzer, 19-7, 19-12 to 19-17, 19-26  
  external, 19-15, 19-26  
  internal, 19-13  
  manual, 19-13, 19-26  
  sequential, 19-13, 19-26  
  terminology for describing, 19-26  
  trigger delay, 19-15, 19-26  
  (*See also* Triggering)

- Trigger qualifier, 19-13, 19-26
- Trigger word, 19-13, 19-26
- Trigger word expander, 19-15
- Trigger word recognizer, 19-13, 19-26
- Triggering:
  - center, 19-16, 19-26
  - delayed, 19-15, 19-26
  - edge, 19-26
  - external, 19-15, 19-26
  - glitch, 19-26
  - internal, 19-13
  - manual, 19-13, 19-26
  - post-, 19-26
  - pretrigger, 19-26
  - sequential, 19-13, 19-26
 (See also Trigger function, logic analyzer)
- Trimmer capacitors, 2-12
- Trimmer variable resistors, 1-14
- Truth table, 14-6
- TTL (transistor-transistor logic or  $T^2L$ )
  - circuits, 15-3 to 15-6
  - high-speed, 15-5
  - low-power, 15-4
  - Schottky-clamped, 15-5
  - standard, 15-3
- Tube rectifiers, 16-6
- Tunable circuits, 10-4
- Tuned-circuit oscillators, 12-30
- Tuned circuits, 10-1 to 10-18
  - bandwidth of, 10-4, 10-15
  - fixed-, 10-3
  - parallel- (see Parallel-tuned circuits)
  - quality factor of, 10-16
  - selectivity of, 10-4, 10-15
  - series- (see Series-tuned circuits)
  - single-, 10-4
  - tunable, 10-4
  - uses of, 10-2
- Tunnel diodes, 8-26
- Turn-off time, 14-19
- Turn-on time, 14-19
- Two-phase to three-phase transformation, 5-33
- Unidirectional bus, 18-3
- Unijunction transistors (UJT), 8-22 to 8-24
  - programmable (PUT), 8-24
- Unit cell, 9-16
- Unit magnetic poles, 4-4, 4-7
- Unity power factor, 10-11
- Up counter, 15-14
- Up-down counter, 15-14
- Vacuum capacitors, 2-9
- Vacuum evaporation, 9-17
- Vacuum-tube voltmeters (VTVM), 7-16 to 7-18
- Vacuum-variable capacitors, 2-11
- Vallari effect, 11-25
- Valley current, 8-23
- Valley voltage, 8-23
- Variable attenuators, 20-22
- Variable autotransformers, 5-24 to 5-26
- Variable capacitors, 2-11 to 2-13, 2-15 to 2-17
- Variable resistors, 1-13 to 1-16
  - carbon, 1-15
  - cermet, 1-15
  - conductive plastic, 1-15
  - contact resistance of, 1-6
  - end resistance of, 1-6
  - ganged, 1-13
  - multiple-turn, 1-13
  - resolution of, 1-6
  - setting stability of, 1-6
  - single-turn, 1-13
  - taper of, 1-5
  - trimmer, 1-14
  - wiper (slider) current in, 1-6
  - wire-wound, 1-10, 1-16
- Varistors, 1-19
- Virtual ground, 13-2
- Volta, Alessandro, 17-1
- Voltage:
  - breakdown, 2-5
  - peak, 8-23
  - saturation, 14-5
  - self-heating, 1-18
  - series-tuned, 10-8 to 10-10
  - source of, 6-2
  - valley, 8-23
  - working, 1-4, 2-5
- Voltage amplifiers, 21-32
- Voltage-controlled current source, 6-2
- Voltage-controlled voltage source (VCVS), 6-2, 22-6
- Voltage dividers, 6-10
- Voltage doublers, 16-9, 16-10
- Voltage feedback, 12-8, 12-28
- Voltage-feedback stabilization, 12-8
- Voltage follower, 13-4
- Voltage gain, 12-3
- Voltage regulation of a transformer, 5-19 to 5-21
- Voltage regulators, 8-8
- Voltage standing wave ratio (VSWR), 20-14
- Voltage-transfer characteristic, 13-18

Voltage triplers, **16-10**  
Voltage-variable capacitors, **2-15 to 2-17**  
Voltmeters, **7-4, 7-16 to 7-18, 7-22**  
Volumetric efficiency, **2-7**

Watt (unit), **6-5**  
Waveguides, **20-5**  
Wavelength, **20-1**  
Weighted-resistor converters, **15-20**

Wheatstone bridge, **7-25 to 7-28**  
Wien bridge, **7-32**  
Wiper (slider) current, **1-6**  
Wire-wound resistors, **1-10**  
Wire-wound variable resistors, **1-16**  
Working voltage, **1-4, 2-5**

Z80 microprocessors, **18-24**  
Zener breakdown, **8-8**  
Zener diodes, **8-8**

## ABOUT THE EDITORS

MILTON KAUFMAN, president of Electronics Writers and Editors, Inc., has extensive experience both as a working engineer and as a writer. He has been a senior engineer with CBS Columbia Color TV Lab, Sperry Gyroscope Co., and Grumman Aircraft (LM-Apollo Projects). He is editor of McGraw-Hill's *Handbook for Chemical Technicians*, *Handbook of Operational Amplifier Circuit Design*, *Handbook of Electronics Calculations for Engineers and Technicians*, and *Handbook of Microcircuit Design and Application*. He has also authored, coauthored, or edited numerous other books, articles, and papers on electronics.

ARTHUR H. SEIDMAN, retired, Professor of Electrical Engineering and Acting Dean, School of Engineering, Pratt Institute, has over ten years' experience in industry, including a position as a senior engineer at Sperry Gyroscope Co. He has authored, coauthored, or edited numerous books, articles, and papers on electronics, including McGraw-Hill's *Handbook of Electronics Calculations for Engineers and Technicians*.



21-61



Electronics

An easy-to-use, portable sourcebook for today's electronics professional

# Electronics Sourcebook FOR TECHNICIANS AND ENGINEERS

Reach for this handy source when you need detailed specifics on components, circuits, power supplies, operational amplifiers, transistors, and much, much more!

Designed to make your day-to-day work easier than ever before, this expertly condensed, softcover edition of the highly-acclaimed *Handbook for Electronics Engineering Technicians*, Second Edition, retains the same helpful, authoritative, in-depth coverage as its famous parent volume. The book you have in your hands is packed with worked-out examples, explanations, illustrations, and practical applications to help engineers and technicians working in electronics solve on-the-job problems quickly and easily.

Here in a single, take-anywhere volume you will find up-to-the-minute coverage of such vital topics as logic analysis and signature analysis . . . microwaves . . . fiber optics . . . active filters . . . digital multimeters . . . microprocessors . . . semiconductors . . . and oscilloscopes. You will also find information about integrated circuit technology, op amps, digital circuits, special types of batteries and their characteristics, component selection, and how to use a wide variety of equipment.

Whether you need a practical on-the-job reference, a useful refresher, an easy-to-comprehend training aid, or a quick update on the latest developments in the field, no other book offers so much valuable information at such a great price.

Also by Kaufman and Seidman:

HANDBOOK OF ELECTRONICS CALCULATIONS  
For Engineers and Technicians  
Second Edition  
864 pages, 725 illustrations

HANDBOOK FOR ELECTRONICS ENGINEERING TECHNICIANS  
Second Edition  
752 pages, 913 illustrations

Cover Design: Richard Adelson  
Cover Photo: © Dori Carroll/The Image Bank



**McGraw-Hill Book Company**  
Serving the Need for Knowledge  
1221 Avenue of the Americas  
New York, NY 10020

ISBN 0-07-033559-1